

# **Linguistische Probleme in der maschinellen Übersetzung**

– Topik und Fokus –

Oh, Young-Hun(Chung-Ang Univ.)

## **1. Einleitung**

Der Bereich der maschinellen Übersetzung befindet sich gegenwärtig in einer Phase der Neuorientierung und des starken Wachstum weltweit. Maschinelle Übersetzung hilft, den Übersetzungsprozess zu beschleunigen und effizienter zu gestalten. Mehr Auftragsvolumen können in weniger Zeit erledigt werden. In diesem Zusammenhang muss klar herausgestellt werden, dass maschinelle Übersetzung den Übersetzerberuf in keinster Weise überflüssig machen wird. Vielmehr ist es eine Technologie, welche der Humanübersetzer als Werkzeug einsetzt, um seinen Arbeitsprozess im Aufwand zu mindern und gleichzeitig qualitativ anzureichern.

Im Abschnitt 2 wird der Begriff von Topik und Fokus aus der linguistischen Sicht eingeführt. Danach wird die Identifizierung von Topik und Fokus dargestellt, wobei die Definition eines Kontextes in einem Experiment vorgestellt wird, und die Regel und Regelverwendung von Topik und Fokus und Salient-Topik eingeführt. Im Abschnitt 3 wird ein Text sich mit der Integration im Generierungssystem beschäftigen, wobei ganz einfache Regeln für die Identifikation von Topik, Salient-Topik und Fokus gegeben sind. Im Abschnitt 4 werden einige Probleme in der Übersetzung behandelt. Abschliessend folgt eine kurze Zusammenfassung.

## 2. Die Identifizierung von Topik und Fokus

Die Wichtigkeit der Kontextfaktoren für einen gesamten Kommunikationsprozess und für Teilprozesse wie die Informationsdistribution und die Markierung von Topik und Fokus ist der Grund dafür, eine Definition *context* für diesen beschränkten Bereich anzustreben. Die Definition soll die traditionell einfachen Fragekriterien ersetzen, welche für die Bestimmung von Topik und Fokus eines Einzelsatzes im Gebrauch sind, und eine echte semantisch-pragmatische Definition von Topik, Salient Topik und Fokus ermöglichen.

### 2.1. Defintion

Fokus:

Grob gesprochen ist der Fokus ein Teil eines Satzes, mit dem im gewissen Sinne etwas Neues mitgeteilt wird; der Rest des Satzes - der Topik - wird als bekannt vorausgesetzt.

Inhaltliche Bestimmung von Fokus:

Bei einer Äusserung eines Satzes S ist jener Teil von S der Fokus  $F_k(S)$ , dessen Funktion in S nicht aufgrund des "relevanten Kontextes"<sup>1)</sup> bekannt ist. (Die übrigen Teile von S<sub>i</sub> bilden das Topik  $T_k(S_i)$ ).

Topik:

Phonologisch ist Topik jene Konstituente, die keinen Satzaccent trägt; stellungssyntaktisch ist Topik die Konstituent, die gegenüber der unmarkierten Wortabfolge vorangestellt ist (also Subjekt oder eben statt des

---

1) Relevanter Kontext(RK) ist das, was man als das gemeinsame Vorwissen von Sprecher und Hörer bezeichnen kann und was Chomsky - in einer ungewöhnlichen Verwendungen des Wortes - Präsuppositionn nennt.

Subjekts eine andere Satzkonstituente); diskurssemantisch ist Topik jene Konstituente, über die gesprochen wird (=Thema) oder die als thematisch zugrundeliegend verstanden wird (=Präsupposition) - den Rest des Satzes nennt man dann COMMENT, KOMMENTAR oder FOKUS. Syntaktisch ist es als Element im Vorfeld zu verstehen.

Dies kann im Deutschen u.a das Subjekt, aber auch jenes andere Satzglied sein. Als Subjekt trägt es ausser in kontrastierende Verwendung keinen Akzent (unmarkierte Stellung); jenes andere Satzglied kann im Vorfeld ebenso wie anderswo Akzentträger sein, also in die Topikrolle verschoben durchaus Rhema-(Fokus)Status haben, kann dort aber ebenso unakzentuiert bleiben.

## 2.2. Die Definition eines Kontextes in einem Experiment

Das folgende Diskursmodell besteht aus der Aktionssequenz TAKE und PUT, die sich als Konstante in den gemachten Konversationen erwiesen hat. Die Anzahl der in der TAKE-Aktion enthaltenen Blöcke bestimmt die Anzahl der nachfolgenden PUT-Aktion. Zuerst wird eine Sprache  $L_1$  definiert, die eine Weltbeschreibung zulässt, und dazu Aussagen wie  $green(b_1)$ ,  $on(b_1, b_2)$  verwendet, z. B.  $UHB(b_1)$  würde bedeuten, dass  $b_1$  ein Element der ungeordneten Menge von Blöcken, die dem Hörer zur Verfügung stehen. Für den Kommunikationsprozess sind die Kenntnisse und Erfahrung von Teilnehmern genau so wichtig wie die greifbaren um sie herumliegenden Dinge.

Deshalb ist eine Sprache  $L_2$  definiert, die erlaubt, von den Annahmen des Sprechers über den Hörer und seine Welt zu sprechen. Tatsächlich können sich der Sprecher und der Hörer in dem Experiment nicht sehen; sie verlassen sich hauptsächlich auf die Annahmen über ihre jeweiligen gegenseitigen Kenntnisse. Aussagen von  $L_2$  sind z. B.  $amk(a_1)$ , wobei  $amk$  ein Eigenschaftssymbol von  $L_2$  ist und  $a_1$  die Aussage von  $on(b_1, b_2)$  ist.

Die gesamte Aussage  $\text{amk}(\text{on}(b_1, b_2))$  heisst, dass der Sprecher die beiderseitige Kenntnis annimmt, dass Block  $b_1$  auf Block  $b_2$  gesetzt ist. Andere Beispiele von Aussagen in der Sprache  $L_2$  sind  $\text{id}(b_1)$ , das bedeutet, dass Block  $b_1$  in einer TAKE-Aktion identifiziert wurde,  $\text{ueq}(\text{Take})$  bedeutet dabei, dass TAKE eindeutig interpretiert werden kann und weiter bedeutet  $\text{apl}(l_1)$ , dass  $l_1$  als eine potenzielle Position für die bewegten Blöcke angenommen wird. Der Kontext  $C$  in diesem Diskursmodell wird als ein Mengenpaar von Aussagen in der Sprache  $L_2$  betrachtet:  $\langle CO_1, CO_2 \rangle$ .  $CO_1$  enthält eine Aussage über die Welt eines Sprechers und Annahmen über die beiderseitigen Kenntnisse, die im Lauf eines Experiments unverändert bleiben.  $CO_2$  beinhaltet die Annahmen eines Sprechers über die Blöcke eines Hörers, z. B. ihre aktuellen und potenziellen Positionen.

### 2.3. Die Operationalisierung der Kategorie-Zuweisung

Die Analyseeinheiten sind die semantischen Repräsentation einer Äusserung aus dieser 'Blocksworld'. Zu jedem Element der semantischen Repräsentation werden semantische Kategorien operational zugewiesen. In dieser Phase wird eine formale Definition der semantischen Repräsentation und der Operationalisierungsregel gegeben. Jede semantische Repräsentation eines illokutionären Plans ist eine geordnete Menge  $IP = \langle x_1, \dots, x_n \rangle$ , wobei  $x_1$  ein Verb ist und die restlichen Elementen  $x_2$  bis  $x_n$  den Elementen des Kasusrahmens des Verbs  $x_1$  entsprechen. Für jedes Element  $x$  von illokutionärem Plan gibt es eine individuelle Konstante in der Sprache  $L_2$ , die  $x^*$  heisst. Die Zuweisung der semantisch-pragmatischen Kategorie zu den Elemente von  $IP$  ist eine Funktion, die jedes Paar  $(x, C)$  auf die semantisch-pragmatische Kategorie von  $x$  abbildet, und den Zustand  $x$  bezüglich  $C$  repräsentiert, wobei  $x \in IP$  und  $C$  der Kontext ist.

Die kontextuellen Werte sind *gegeben(g)*, *gewählt(ch)*, *ausgesagt(m)*, *erwähnt in dem vorangegangenen Satz(mp)* und *ihre Negation -ch, -m*,

-mp. -g kommt nicht vor.

Diese Symbole bilden das Alphabet  $A = \{ch, m, mp, -ch, -m, -mp\}$ . Die Operationalisierungskriterien sind folgendermassen:

- (1) Falls  $x \in IP$  ist, dann gilt:
  - (i) falls es eine Eigenschaft  $y$  von  $L_1$  gibt, die besagt, dass  $y(x^*) \in CO_1 \cup CO_2$  und dass jedes andere Objekt  $x' : y(x') \notin CO_1 \cup CO_2$ , dann  $g(x)$ . Dieses Kriterium wird in diesem Fall angewendet, wo es nur ein  $X^*$  gibt, das die Eigenschaft Hörer( $x^*$ ) hat.
  - (ii) falls  $amk(x^*)$  und  $ueq(x^*) \in CO_1 \cup CO_2$ , dann  $g(x)$ . Dieses Kriterium wird angewendet für  $x = TAKE$  als ein Element der Aktionsfolge  $\langle TAKE, PUT \rangle$ , die als beiderseitige Kenntnis betrachtet wird und für den Hörer eindeutig interpretierbar ist.
- (2) Falls  $x \in IP$ ,  $UHB(x^*) \in CO_2$ , dann  $ch(x)$  und  $-m(x)$ . Dieses Kriterium wird z. B. angewandt auf die Elemente der ungeordneten Menge von Blöcken des Hörers.
- (3) Falls  $x \in IP$ ,  $id(x^*) \in CO_2$ , dann gelten die folgende Bedingung:
  - (i) falls  $x$  das erste Objekt in einer PUT-Aktionsfolge ist, dann  $ch(x)$ ,  $m(x)$ ,  $mp(x)$ .
  - (ii) falls  $x$  ist weder das erste noch das letzte Objekt in einer PUT-Aktionsfolge ist, dann  $ch(x)$ ,  $m(x)$  und  $-mp(x)$ .
  - (iii) falls  $x$  das letzte Objekt in einer PUT-Aktionsfolge ist, dann  $-ch(x)$ ,  $m(x)$  und  $mp(x)$ .
  - (iv) falls  $x$  das einzige Objekt in einer PUT-Aktionsfolge ist, dann  $-ch(x)$ ,  $m(x)$  und  $-mp(x)$ .
- (4) Falls  $x \in IP$  und  $apl(x^*) \in CO_2$ , dann  $ch(x)$  und  $-m(x)$ .

Dieses Kriterium wird angewendet, wenn der Sprecher annimmt, dass es eine Position  $on(b_1)$  gibt, unter anderen, die sich potenziell von dem gerade bewegten Block besetzen lässt.

## 2.4. Die Regel von Topik(t), Salient Topik(st) und Fokus(f)

Die Regel von Topik:

- (5) If  $g(x) \in IP$  then  $t(x)$ .
- (6) If  $(-ch, m, mp(x)) \in IP$  then  $t(x)$ .

Die Regel von Salient Topik:

- (7) If  $(ch, m, mp(x)) \in IP$  then  $st(x)$ .
- (8) If  $(ch, m, -mp(x)) \in IP$  then  $st(x)$ .
- (9) If  $(-ch, m, -mp(x)) \in IP$  then  $st(x)$ .

Die Regeln (7) und (8) können durch die äquivalente Regel ersetzt werden, d. h.

- (7\*) If  $(ch, m(x)) \in IP$  then  $st(x)$ .

Die Regel von Fokus:

- (10) If  $(ch, -m(x)) \in IP$  then  $f(x)$ .<sup>2)</sup>

## 2.5. Übersetzung

Um eine Vorstellung über die Relation zwischen den Repräsentationsargumenten von illokutionären Plan und ihren kontextuellen Zuständen zu geben, wird das Beispiel 1 in der folgenden Reihe dargestellt: 'rule number', 'assigned category' and 'contextual information'.

Beispiel 1:

Die Argumente des illokutionären Plans (TAKE, ADRESSEE, OBJECT) bringen die folgenden Wertzuschreibungen mit:

- (1 ii) ---->  $g(\text{TAKE})$

---

2) Pignataro, Vincenza: A computational Approach to Topic and Focus in a Production Model. Coling 88. S. 516.

- $\text{amk}(\text{TAKE}^*) \in \text{CO}_1, \text{ueq}(\text{TAKE}^*) \in \text{CO}_1$
- (1 i) ---->  $g(\text{ADRESSEE})$   
 $\text{hearer}(\text{ADRESSEE}^*) \in \text{CO}_1$
- (2) ---->  $((\text{ch}, -\text{m}) (\text{OBJECT}))$   
 $\text{UHB}(\text{OBJECT}^*) \in \text{CO}_2$

Deshalb lautet der neue IP'  $\langle g(\text{TAKE}), g(\text{ADRESSEE}), ((\text{ch}, -\text{m})(\text{OBJECT})) \rangle$  und die Anwendung von Regel(5) auf das erste und zweite Argument und von Regel(10) auf das dritte Argument von IP' ergibt folgende Umwandlung: IP':  $\langle t(\text{TAKE}), t(\text{ADRESSEE}), f(\text{OBJECT}) \rangle$ .

Die Oberflächenstruktur des illokutinären Plans ist 'du nimmst einen roten Klotz'. Dieser Satz wird als '(you) take a red block' übersetzt.

Beispiel 2:

Die Regelanwendung von IP =  $\langle \text{PUT}, \text{ADRESSEE}, \text{OBJECT}, \text{GOAL} \rangle$ :

- (1 ii) ---->  $g(\text{PUT}),$   
 (1 i) ---->  $g(\text{ADRESSEE}),$   
 (3 iv) ---->  $((-\text{ch}, \text{m}, \text{mp}) (\text{OBJECT})),$   
 (4) ---->  $((\text{ch}, -\text{m}) (\text{GOAL})).$

In diesem Fall ist der neue illokutionäre Plan IP':  $\langle g(\text{PUT}), g(\text{ADRESSEE}), ((-\text{ch}, \text{m}, \text{mp})(\text{OBJECT})), ((\text{ch}, -\text{m}) (\text{GOAL})) \rangle$ .

Die Regelanwendung (5) auf das erste und zweite Argument, die Regelanwendung (6) auf das dritte und die Regelanwendung (10) auf das vierte Argument von IP' ergibt:

IP'':  $\langle t(\text{PUT}), t(\text{ADRESSEE}), t(\text{OBJECT}), f(\text{GOAL}) \rangle$ .

Die Oberflächenstruktur 'du stellst ihn auf den Tisch' wird als 'put it on the table' übersetzt.

Beispiel 3:

Um die Anwendung von Salient-Topik-Regel darzustellen, nehmen wir an,

dass die folgende Äusserung als Folge eines illokutionären PUT-Plans entsteht: 'du nimmst einen roten und einen blauen Klotz' wird als 'take a red and a blau block' übersetzt.

Zwei illokutionäre PUT-Pläne folgen:

$IP_1 = \langle \text{PUT}, \text{ADRESSEE}, \text{OBJECT}_1, \text{GOAL} \rangle,$

$IP_2 = \langle \text{PUT}, \text{ADRESSEE}, \text{OBJECT}_2, \text{GOAL} \rangle.$

Für das erste, zweite und vierte Argument von der Menge  $IP_1$  und  $IP_2$  gilt die gleiche Beschränkung wie im oben PUT-Beispiel. Für das dritte Argument wird die folgende Regel angewendet:

(3 i) ----> ((ch, m, mp) (OBJECT<sub>1</sub>)),

(3 iii) ----> ((-ch, m, -mp) (OBJECT<sub>2</sub>)).

Deswegen sind die neuen illokutionären Pläne:

$IP_1' = \langle g(\text{PUT}), g(\text{ADRESSEE}), ((\text{ch}, \text{m}, \text{mp}) (\text{OBJECT}_1)), ((\text{ch}, -\text{m}) (\text{GOAL})) \rangle,$

$IP_2' = \langle g(\text{PUT}), g(\text{ADRESSEE}), ((\text{ch}, \text{m}, \text{mp}) (\text{OBJECT}_2)), ((\text{ch}, -\text{m}) (\text{GOAL})) \rangle.$

Die Regelanwendung (5) auf das erste und zweite Argument, die Regelanwendung (7) auf das dritte Argument in  $IP_1'$ , die Regelanwendung (9) auf das dritte Argument von  $IP_2'$  und die Regelanwendung (10) auf das vierte Argument von  $IP_1'$  und  $IP_2'$  ergibt:

$IP_1'' = \langle t(\text{PUT}), t(\text{ADRESSEE}), st(\text{OBJECT}_1), f(\text{GOAL}) \rangle,$

$IP_2'' = \langle t(\text{PUT}), t(\text{ADRESSEE}), st(\text{OBJECT}_2), f(\text{GOAL}) \rangle.$

Die Oberflächenstruktur 'du stellst den roten auf den grünen und den blauen auf den roten' wird als 'put the red on the green and the blue on the red' übersetzt.

Falls das dritte Argument in einem illokutionären TAKE-Plan aus einer Liste von mehreren Objekten besteht, dann gilt  $ch(\text{OBJ})$  und  $-m(\text{OBJ})$  für jedes Objekt OBJ.

Dies lässt sich durch den Ausdruck  $(\text{ch}, -\text{m})^*$  der formalen Sprache über das Alphabet A abkürzen. Für jedes dritte Argument des illokutionären

PUT-Plans gilt das folgende: für das erste Argument (ch, m, mp), für das Objekt  $2$  zu  $n_1$  (ch, m, -mp) und für das letzte Objekt (-ch, m, -mp). Dies kann durch den Ausdruck (ch, m, mp), (ch, m, -mp)\*, (-ch, m, -mp) abgekürzt werden. Wenn der Satz 'du nimmst zwei/drei grüne Klötze' bedeutet 'take three green blocks', dann handelt es sich um eine kohäsive nachfolgende Äusserung, d.h. 'du stellst einen ..., einen... und einen/den letzten ...' bedeutet 'put one ..., one ... and one/the last one ...'. Wenn die genommene Blöcke 'two reds and a blue' sind, muss die nachfolgende Antwort sein: 'du stellst einen roten ..., einen roten ... und den blauen ...': das bedeutet 'put one red ..., one red ... and the blue ...'.

### 3. Die Integration im Generierungssystem

Die Kontrolle der Konversationsdynamik durch die Werte ausgesagt(m), erwähnt in der vorangegangenen Äusserung (mp) und die Markierung der Äusserungselemente Topik(t) und Fokus(f) sind nur zwei aus den verschiedenen Subprozesse, die im Lauf des Hauptproduktionsprozesses parallel ablaufen. In der automatischen Produktion der natürlichen Sprache, sowohl gesprochener Sprache wie auch geschriebener Sprache, müssen die thematische Progression einer Äusserungsfolge und ihre formale Kohäsion mitberücksichtigt werden. Die Regeln für die Identifikation von Topik(t), Salient Topik(st) und Fokus(f) garantieren die Kohärenz des thematischen Fortschritts von zwei oder mehrere nachfolgenden Äusserungen der Aktionsfolge. Zwei ganz einfache Regeln für den thematischen Fortschritt sind mit der jeweiligen Nummer von obigen Beispielen im folgenden gegeben.

R1: Das einzige fokussierte OBJEKT einer TAKE-Aktion wird zu dem topikalisierten Objekt der nachfolgenden PUT-Aktion (Beispiel 1 und 2).

R2: Die zwei/drei fokussierten Objekten der TAKE-Aktion werden zu dem OBJEKT der nachfolgenden zwei/drei PUT-Aktion und werden als ein Salient-Topik bezeichnet (Beispiel 3).

Die Topik, Salient Topik und Fokus-Identifikationsregeln erlauben es, Vorhersagen über die Akzentdistribution zu machen. "Tatsächlich werden die Elemente, die Salient-Topik und Fokus betitelt sind, akzentuiert. Das Topik-Element erhält keinen Akzent. In dieser Phase der Arbeit werden allen Argumenten der Propositionen Akzente zugewiesen."<sup>3)</sup> Die Akzentzuweisung zum Adjektiv anstatt zum Nomen in Phrasen wie '...den roten...' schliesst die Anwendung der gleichen Kriterien ein. Zur Generierung kohäsiver Oberflächenstrukturen ist es auch vonnöten, zu wissen, wann eine definite Artikel innerhalb von Nomenphrasen oder ein indefinites Pronomen verwendet wird. Diese Wahl hängt von der pragmatischen Entscheidung ab, einen oder mehrere Blöcke aufzunehmen und von den Eigenschaften, die die in Frage kommenden Objekte gemeinsam haben.

## 4. Die Beschreibung von Problemen

### 4.1 Der Negationsskopus

"(11) The eec is not controversial, because it is a multi-national ORGANIZATION.

(12) Because it is a mult-national organization, the eec is not CONTROVERSIAL."<sup>4)</sup>

---

3) Vgl. Pignataro, V.: A Computational Approach to topic and focus in a production model. Coling 88. S. 517.

4) Steiner, E.H., Winter-Thielen, J.: On the semantics of focus phenomena in EUROTRA. Coling 88. S. 630.

Die Beispiele (11) und (12) sind identisch, ausser für die Position des Adverbial-Satzes, und für die Zuweisung der primären Information; FOKUS, der durch die Hauptbetonung realisiert wird, wird durch Grossbuchstaben symbolisiert. Aber ihre semantische Interpretation unterscheidet sich hinsichtlich des Negationsskopus: das Beispiel (11) ist uneindeutig im Bezug auf die Frage, ob die in dem Matrix-Satz zum Ausdruck gebrachte Proposition wahr ist oder nicht. Vergleiche (13):

- (13) The eec is not controversial, because it is a multi-national ORGANIZATION.  
It is controversial because it is EXPENSIVE.<sup>5)</sup>

Der zweite Satz im Beispiel (13) ist nicht möglich mit dem Beispiel (11), während es mit dem Beispiel (12) möglich ist. Im Beispiel (11) schliesst der Negationsskopus den Adverbialsatz ein, während das Negationsskopus im Beispiel (12) den Adverbialsatz nicht einschliesst.

Im allgemeinen scheint es zu sein, dass das Negationsskopus wohl alles in der *Clause* rechts der Negationspartikel einschliesst, und auch die Konstituente, die ein lexikalisches Item mit FOKUS beinhaltet. Deshalb ist der Wahrheitswert von den Beispielen (11) und (12) verschieden. Das impliziert, dass beide keine akzeptablen Übersetzungen voneinander in der am Wahrheitswert orientierten Semantik sind.

## 4.2 Der Skopus von Quantor

- (14) Everyone in this room speaks two languages.  
(15) Two languages are spoken by everyone in this room.<sup>6)</sup>

---

5) Steiner, E.H., Winter-Thielen, J.: Ebd. S. 630.

6) Hajičová, E., Sagall, P., Skoumalová, H.: An Automatic Procedure for Topic-Focus Identification. In: Association for computational Linguistics. 1995. S. 82.

Diese Beispielsätze sind ein klassisches Beispiel, das in der Literatur immer wieder diskutiert worden ist. Im Fall vom Beispiel (14) spricht jede Person im betroffenen Raum mindestens zwei Sprachen. Aber es ist nicht gesagt, dass jeder die gleichen zwei Sprachen spricht. Im Fall vom Beispiel (15) versteht jeder im Raum die gleichen zwei Sprachen. Die Schwierigkeit scheint zu sein, dass der Typ und das Skopus des zugrundeliegenden Quantors je nach der linearen Folge sich ändern. Hier unterscheiden sich die Beispiele (14) und (15) bereits auf der Ebene der D-Struktur, welche als (14') bzw. (15') anzusetzen ist.

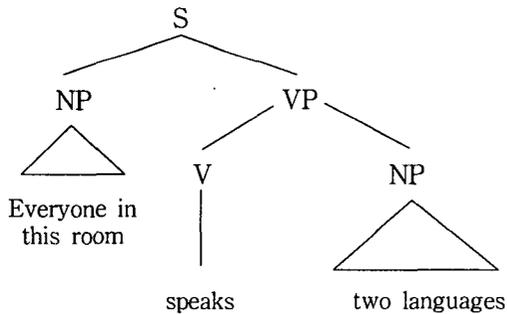


Abbildung 14'

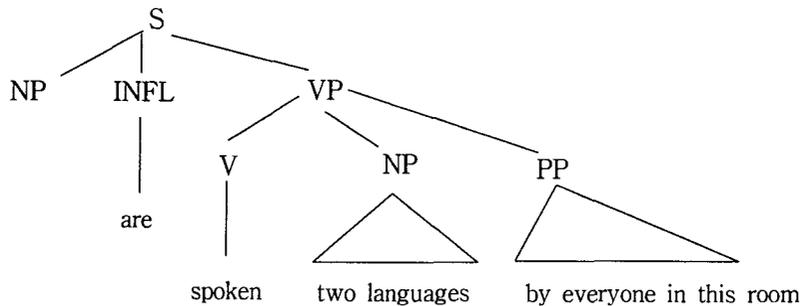


Abbildung 15'

Diese Probleme kommen sogar in Sprachen mit freierer Wortfolge als im Deutschen vor.

- (16) Einige von uns haben eine Sorge. ----> Some of us have a worry.
- (17) Eine Sorge haben einige von uns. ----> A Worry some of us have.
- (18) Wir alle arbeiten an einer Lösung. ----> All of us a working towards a solution.
- (19) An einer Lösung wird von allen von uns gearbeitet. ----> A solution is worked on by all of us.<sup>7)</sup>

In den Beispielen (16) und (18) beinhaltet die indefinite Artikel in der Objekt NP nicht notwendigerweise ein und das gleiche für alle Subjekten, während die Implikation in den Beispielen (17) und (19) gegeben ist. Es sei angemerkt, dass das Beispiel (19) eine Passivierung enthält, während das Beispiel (17) Thematisierung ohne Passivierung einschliesst.

Dies zeigt, dass es hier nicht um GENUS geht, sondern um FOKUS. Es scheint aber eher Typ als Skopus von Quantor zu sein. Zudem wird die Beurteilung der *nativen Speaker* durch die Intonation und andere Faktoren beeinflusst.

### 4.3 Akzeptable Frage-Antwort-Paare

Für die Illustration von diesem Problem sehen wir die Beispiele (20) bis (25) an:

- “(20) When did Parliament decide on the budget?
- (21) Parliament decided on the budget last WEEK.
- (22) It was last WEEK that Parliament decided on the budget.
- (23) ? last week parliament decided on the BUDGET.

---

7) Steiner, E.H., Winter-Thielen, J.(1988): A,a.O, S. 631.

(24) ? It was on the BUDGET that Parliament decided last week.

(25) ? What Parliament DID last week was decide on the budget."<sup>8)</sup>

Die Beispiele (21) und (22) sind als eine Antwort auf das Beispiel (10) geeignet, aber die Beispiele (23) - (25) sind zumindest fraglich. Man beobachtet, dass das Thematisierungsmittel wie in den Beispielen (22) - (25) nicht in einfachem syntaktischen Transfer zwischen verschiedenen europäischen Sprachen übersetzt wird.

Die deutsche Sprache z. B. kann sich einfach die lineare Wortstellung verändern, wo das Englische andere Mittel benötigt (z. B. Clefting, Pseudo-Clefting<sup>9)</sup>). Weiter ist zu beobachten, dass der Fokus mit dem thematisierten Konstituenten im Fall der markierten Thema-Zuweisung zusammentrifft, während der Rest des Satzes als Präsupposition fungiert. Dies ist von Belang für die geschriebene Sprache, wo keine direkte Enkodierung der Intonation möglich ist. Im Fall der WH-Frage läuft der Fokus allgemein mit dem WH-Element ab. Im Fall der unmarkierten Thema-Zuweisung fällt der Fokus im Englischen auf das letzte vollständige lexikalische Item eines Satzes.

Zurück zur Frage der Satzübersetzung mit bestimmten Typen markierter Themen.

Vergleiche die Beispiele (26) bis (29).

---

8) Steiner, E. H., Winter-Thielen, J.(1988): Ebd.

9) Zu Hervorhebungszwecken in einem Satz der Form es ist + NP vorausgezogener Nominalteil der nachfolgenden Fügung; Es ist dieser Mann, den ich meine; transformationell aus einem einzigen Satz durch Linksversetzung abgeleitet (daher der Name). Im Gegensatz dazu denkt man sich die Erscheinung des Pseudo-cleft-sentence (Sperrsatzes), wen ich meine, ist dieser Mann, als transformationelle Rechtsversetzung abgeleitet. Hinsichtlich der Funktion im Textablauf unterscheiden sich die beiden Konstruktion nicht voneinander: beide schaffen zu semantischen Kontrazwecken syntaktisch eine eindeutig markierte Fokuskonstituente. Abraham, Werner: Terminologie zur neueren Linguistik, Max Niemeyer Verlag Tübingen, 1988, S. 771.

- “(26) Zu einem grossen UNGLÜCK wurde die Entscheidung erst durch die weitere ENTWICKLUNG.
- (27) ? A big DESASTER the decision only became of further DEVELOPMENTS.
- (28) It was only through further DEVELOPMENTS that the decision became a DESASTER.
- (29) ? Es war nur durch die weitere Entwicklung, dass die Entscheidung zu einem Unglück wurde.”<sup>10)</sup>

Das Beispiel (27) ist eine fragliche Übersetzung von dem Beispiel (26), aber das vorhandene Framework diktiert entweder, oder dies erzeugt die kanonische Form, die dem Beispiel (30) als einer Übersetzung von dem Beispiel (26) zugrundliegt:

- (30) The decision became a desaster because of further DEVELOPMENTS.

Im Beispiel (30) fehlt allerdings der Doppel-Fokus vom Beispiel (26) und ist daher problematisch. Auf der anderen Seite ist das Beispiel (29) eine fragliche Übersetzung vom Beispiel (28), denn das syntaktische Mittel des *Clefting* führt nicht vom Englischen ins Deutsche. Es ist einleuchtend, dass

- diese Probleme nicht in isolierten Beispielen vorkommen, sondern mit Thematisierungsmittel im allgemeinen.
- diese Probleme sich zwischen Sprachen sehr vergrössern, die weniger gemeinsam als das Englische und das Deutsche haben.

## 5. Schlussbemerkungen

Syntax und Semantik konzentrieren sich in erster Linie auf den Satz als

---

10) Steiner, E.H., Winter-Thielen, J.(1988): A.a.O. S. 632.

Gegenstand ihrer sprachwissenschaftlichen Untersuchung. Sätze treten aber normalerweise nicht einzeln auf, sondern werden im Kontext mit anderen Sätzen verwendet, mit denen zusammen sie aufgrund einer Reihe linguistischer und ausserlinguistischer Faktoren einen Text konstituieren. Die Textlinguistik beschäftigt sich mit den allgemeinen Prinzipien und Merkmalen, die eine Folge von Sätzen für einen Rezipienten als einen Text erscheinen lässt. Die Diskursanalyse, die sich mit satzübergreifenden Phänomenen befasst, betont in stärkerem Masse Fragen der kommunikativen Funktion von Texten und schliesst auch dialogische Texte in ihre Untersuchungen mit ein. Entscheidend dafür, dass eine Folge von Sätzen als Text wahrgenommen wird, also in einer bestimmten kommunikativen Funktion wirksam werden kann. Auch für die maschinelle Übersetzung spielt die linguistische Pragmatik entscheidende Rolle.

In der ersten Phase dieser Arbeit wird ein Diskursmodell in einem Experiment eingeführt, wobei die Definition des Kontextes, die Operationalisierungskriterien und die Regel von Topik, Salient-Topik und Fokus, die hier für die Übersetzung angewendet werden, vorgestellt werden. Danach werden einige Übersetzungsbeispiele vorgelegt.

In der zweiten Phase werden einige Probleme aus der linguistischen Sicht dargestellt. Z. B. wie beeinflusst die Position von Negationsskopos die maschinelle Übersetzung in einem Satz? usw.

Eine Untersuchung eines bestimmten Bereich der maschinellen Übersetzung kann im Prinzip nie als abgeschlossen gelten. Die vorliegende Arbeit zu den Topik - und Fokus-Phänomenen bildet hierzu keine Ausnahme.

### **Literaturverzeichnis**

- Abraham, Werner: Terminologie zur neueren Linguistik, Max Niemeyer Verlag Tübingen, 1988.  
Alonso, J.A.: A Model for Transfer Control in the METAL. Proc. TMI 1990,

- Austin.
- Baart, J.L.: Focus and Accent in a Dutch text-to-speech system. 4EACL.
- Brown, P. et al.: Analysis, Statistical transfer, and Synthesis in machine Translation. In: TMI 92.
- Brustkern, Jan: Maschinenlesbare Lexika für die maschinelle Sprachverarbeitung (Repräsentation und Wiederverarbeitung). Holos Reihe Linguistik, Band 3, Bonn 1992.
- Hajicova, E., Sgall, P., Skoumalova, H.: Identifying Topic and Focus by an automatic procedure. 6EACL.
- Okumura, A., Hovy, E.: Lexikon-to-Ontology Concept Association Using a Bilingual Dictionary. AMTA94.
- Pignataro, Vincenza: A computational Approach to Topic and Focus in a Production Model. Coling 88.
- Pyka, Claudius et. al.: Untersuchung über Trends und Perspektiven in der rechnergestützten und maschinellen Übersetzung, Version 0.9. SCS Informationstechniker GmbH. 1990.
- Stechow, Armin von, Sternefeld, Wolfgang: Bausteine syntaktischen Wissens, Westdeutscher Verlag, 1988.
- Steedman, Mark: Surface Structure, Intonation and Focus. In: Natural Language and Speech. Klein, E., Veltman, F.(eds.), Springer Verlag Berlin 1991. S. 21-38.
- Steiner, E.H., Winter-Thielen, J.: On the semantics of focus phenomena in EUROTRA. Coling 88.
- Syea, Anand and Allegranza, Valerio: Clefting and Topicalization, Eurotra Reference Manual, 7.0, 1990.
- Trujillo, A.: Locations in the machine Translation of prepositional Phrase. In: Proc TMI 92.
- Vlk, T.: Topic/focus articulation and intensional logic. Coling 88.
- Yamada, A., Nishida, T., DOshita, S.: Figuring out most plausible interpretation from spatial descriptions. Coling 88.
- Ward, N.: An alternativ to Deep Case for Representing relational Information. Coling 92.
- Winter-Thielen, J, Steiner, E. H.: On the semantic of focus phenomena in EUROTRA. Coling 88.

## 국문요약

### 기계번역에 있어서 언어학적인 문제점

—주제어와 초점어를 중심으로—

오영훈(중앙대학교)

오늘날 기계번역 Maschinelle Übersetzung은 가속적인 발전선상에 놓여있다. 지난 10년 간 컴퓨터로 영어를 타국어로 번역하는 수준은 괄목할 만하다.

본 논문은 기계번역에 있어서 주제어 Topik 및 초점어 Fokus를 중심으로 발생하는 언어학적인 문제점, 특히 의미론적인 문제점을 다루었다. 이를 위해 먼저 주제어와 초점어에 대한 언어학적인 개념을 다루어 보았다. 주제어란 한 문장에서 이미 알려진 사항, 즉 이미 주어진 단어이고, 초점어는 한 문장에서 새로운 사항, 즉 지금 전달하고자 하는 내용을 뜻한다. 제 3장에서는 주제어와 초점어를 번역하기 위해 생성된 규칙들에 근거한 담화모델 Diskursmodell을 살펴보았다. 제 4장에서는 문장을 번역하는데 있어서 의미론상 발생하는 문제점들을 다루었다. 그 문제점들은 다음과 같은 3가지로 요약될 수 있다: 첫째, 문장에서 부정형이 어디에 위치하느냐에 따라 문장의 의미가 달라진다. 둘째, 양화사 Quantor의 형태에 따라 문장의 의미가 달라진다. 셋째, 의문문과 화답문 Antwortsatz 에 있어서 어느 내용을 강조하느냐에 따라 문장의 의미가 달라진다. 예를 들어 독일어는 단순히 단어의 위치만 변화시킬 수 있지만 다른 유럽어나 영어는 다른 방법들이 필요한 셈이다.

본 논문에서 제시되고 있는 기계번역의 규칙들은 주제어와 초점어에 관계되는 한 영어와 독일어에 한정되어 제시되었지만, 향후 한국어와 독일어의 기계번역을 위해 밑거름이 되리라 생각한다.

**[검색어]** 기계번역, 주제어, 초점어, 담화모델  
Maschinelle Übersetzung, Topik, Fokus, Diskursmodell.

오영훈

156-756

서울시 동작구 흑석동 221

중앙대학교 문과대학 독어독문학과

ohy10106@hotmail.com