

Automatisk översättning – en introduktion

Robert Eklund
Stockholms universitet
1996

1 Introduktion

Automatisk översättning har sedan slutet av femtiotalet varit ett av datorlingvistikens huvudområden, även om området som sådant haft sina 'ups and downs'. I denna uppsats skall området maskinöversättning, ofta förkortat MT, presenteras. De olika metoder och filosofier som finns inom området skall presenteras och redogöras för, samt några av de större problem som området har att bjuda på samt hur man på olika sätt försökt lösa dessa.

Större delen av de artiklar och verk som nämns utgör del av det kursmaterial som Louisa Sadler delade ut under NorFa-kursen *Komputationell Semantik och Automatisk Översättning* som gavs i Bergen, 12–16 april 1994. Utöver själva kurslitteraturen har även annan litteratur använts. Av denna vill jag här nämna framför allt två huvudkällor: Arnold et al (1994): *Machine Translation – An Introductory Guide* och Frank van Eynde (1993): *Machine Translation and Linguistic Motivation*. Övrig litteratur refereras till i själva uppsatsen.

2 En kort historik

Ursprunget till själva idén att automatiskt översätta mellan två naturliga språk är givetvis svår att spåra, men följande exempel brukar anges som det tidigaste: Med avstamp i sin korrespondens med den engelske kristallografen Andrew Booth skrev Warren Weaver vid Rockefeller Foundation 1949 ett memorandum där följande två meningar ingick:

I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is to strip off the code in order to retrieve the information contained in the text.

Man hade med viss framgång använt 'datorer' vid inom kryptografi under andra världskriget, och Weavers memorandum väckte ett visst intresse. Redan vid början av 1950-talet arbetade grupper i USA och Europa med problemet, starkt uppbackade ekonomiskt. Även om man nådde viss framgång, och kanske framför allt blev medveten om många frågor som fortfarande är centrala, så började tvivel dyka upp rörande områdets framkomlighet.

De teoretiska tvivlen uttrycktes in en rapport från 1959 där filosofen Bar-Hillel argumenterade att *fully automatic, high quality machine translation* (förkortat FAHQMT) var omöjligt, inte bara för tillfället, utan i princip. Bar-Hillel påpekade med ett antal exempel att språktolkning inte bara har med språk att göra, utan med generell världskunskap, och att denna kunskap inte kan byggas in i en dator. Denna rapport fick stor betydelse eftersom den klargjorde att forskningsområdet berörde fundamentala frågor i den mänskliga språkprocessningen.

Även om det möjligen är så att obegränsad, generell FAHQMT är omöjlig så betyder inte det att *all* MT är omöjlig. År 1964 skapades The Automatic Language Processing Advisory Committee, ALPAC, av The US National Academy of Science för att undersöka kvaliteten, kostnaden och framtidsutsikterna för området. Rapporten var mycket kritisk och ledde till att nästan alla anslag drogs in. Under tio år utfördes den större delen av forskningen inom MT av grupper som bekostades av den amerikanska mormonkyrkan som hade intresse av bibelöversättningar, samt en liten mängd grupper i Kanada, som t ex TAUM i Montreal, som utvecklade METEO (mer därom senare). Även i Europa bedrevs viss forskning.

I stället för att ge forskningsanslag till MT satsades i stället pengar på med grundläggande forskning inom datorlingvistik och artificiell intelligens. Runt 1970 upplevde dock MT en renässans, och sedan dess har intresset och det satsade kapitalet sakta men säkert ökat.

3 Varför MT?

Som Bar-Hillel påpekade är det måhända omöjligt att nå obegränsad, generell FAHQMT. Detta behöver som sagt inte utesluta att MT kan vara av intresse helt och hållet. Först och främst är det möjligt att få MT-system att fungera bra inom avgränsade domäner, där själva kontexten är så begränsad att många eller de flesta av (de centrala) världskunskapsproblemen undviks. Sålunda översätter METEO i Kanada 45000 ord per dag mellan franska och engelska på ett i det närmaste felfritt sätt. Domänen är här väderleksrapporter.

Även om FAHQMT inte är tillgängligt så kan en maskinellt delvis översatt text vara till hjälp. Tillgången på mänskliga översättare är begränsad samt dyr. Om man lyckas klara av bara en liten del av översättningen automatiskt så har man sparat stora summor pengar. Man har uppskattat att av de löpande kostnaderna inom EU står 40-45 % för språkkostnader.

MT är även av intresse rent vetenskapligt. Man har här en testbänk för idéer inom datorlingvistik, artificiell intelligens och lingvistik. Sålunda kan nämnas att programmeringsspråket prolog ursprungligen skapades med MT i åtanke. Även på ett högre filosofiskt plan är MT av intresse eftersom det är ett område som täcker in snart sagt varje form av mänsklig kunskap, och att därför varje framsteg innebär ökad kunskap inom en mängd områden.

4 Vad krävs för MT?

Som Bar-Hillel påpekade har språktolkning i hög grad att göra med kunskap annan än rent språklig. Om man trots detta ämnar försöka skapa ett automatiskt översättarsystem

är det givetvis av intresse att se vilket slags kunskap mänskliga översättare använder sig av. Man brukar räkna med åtminstone fem slags kunskapsområden:

- Kunskap om källspråket (SL).
- Kunskap om målspråket (TL).
- Kunskap om korrespondenser mellan SL och TL.
- Kunskap om världen generellt och om det ämne aktuell text berör specifikt.
- Kunskap om kulturella och sociala konventioner, sedvänjor och dylikt som talaren av SL och TL har.

De sista två punkterna här motsvarar ganska väl Bar-Hillels åsikt om varför FAHQMT är omöjligt, och ingen har heller lyckats förslå en lösning på problemet hur detta slags kunskap skall representeras.

De första tre punkterna, emellertid, kan eventuellt implementeras och motsvarar i 'renare' form lingvistisk kunskap. Den kunskap som här krävs är följande:

- Fonologisk kunskap, dvs språks ljudsystem.
- Ortografisk kunskap, dvs hur språks ljudsystem uttrycks i skrift.
- Morfologisk kunskap, dvs ordbildning i de aktuella språken.
- Syntaktisk kunskap, dvs hur fraser och satser kan byggas upp i de aktuella språken.
- Semantisk kunskap, dvs kunskap om hur olika betydelser uttrycks.

Även om ovanstående lingvistiska problem inte är löst så är de åtminstone åtkomliga. Det existerar således en uppsjö lingvistiska modeller med syfte att omfatta och representera ovanstående delområden.

5 Olika problem vid översättning

Även om vi ovan just sett att MT inom begränsade domäner är görligt, och att den strikt lingvistiska kunskapen i någon mån låter sig fångas och beskrivas i modeller så är det därmed inte sagt att problemet MT är löst. Själva bristen på fungerande system – ett par-tre stycken – talar ju sitt tydliga språk. MT är ett område med stora problem att lösa, även om man enbart ser på språkliga problem och helt undviker tolkningar som kräver kunskap om världen. Vilka är då dessa problem?

Det första problem som varje person som lär sig ett annat språk stöter på är att språk endast undantagsvis uppvisar ett ett-till-ett-förhållande sinsemellan. Dessa så kallade 'missmatchningar' dyker upp på samtliga nivåer, och härnedan skall redogöras för några exempel på hur dessa kan yttra sig.

5.1 "Missmatchning" på olika nivåer

Ett uppenbart problem är att språk inte delar in eller beskriver, "mappar", världen på samma sätt vad gäller ordförråd. Denna missmatchning kan ske på olika sätt och nivåer.

5.1.1 Ord–ord

Redan på ordnivå är det ofta inte uppenbart hur man skall översätta vissa begrepp. ord används ju till att kategorisera omvärldens koncept och fenomen, och olika språk uppvisar olika sätt att dela in världen.

Om vi jämför följande exempel från engelska och japanska/tyska ser vi att översättning kräver mer än bara överföring av ett ord till ett annat:

ENGELSKA	JAPANSKA
<i>wear</i>	<i>haku</i> (på fötterna, benen) <i>kakeru</i> (glasögon) <i>kaburu</i> (på huvudet) <i>hameru</i> (på händerna) <i>shimeru</i> (runt halsen) <i>haoru</i> (rock, jacka) <i>kiro</i> (generiskt) <i>tsukeru</i> (broscher o dyl)

ENGELSKA	TYSKA
<i>wall</i>	<i>Wand</i> (innervägg) <i>Mauer</i> (yttervägg)
<i>eat</i>	<i>essen</i> (människor) <i>fressen</i> (djur)

Den här sortens exempel står att finna mellan alla språk, och exemplen skulle lika gärna kunnat gälla i andra riktningen.

5.1.2 Ord–fras

Ett ord i ett språk kan också motsvaras av en hel fras i ett annat språk. Följande exempel visar hur det engelska *stabbed* motsvaras av den spanska frasen *dio una puñalada a*.

ENGELSKA	SPANSKA
<i>John stabbed Mary</i>	<i>Juan le dio una puñalada a Maria</i>

Detta utgör alltså skillnaden mellan att *kniva någon* och att *utdela ett knivhugg i någon.*, lite odiamatiskt översatt.

5.1.3 Tempus och aspekt

Ytterligare ett problem är när det visserligen är klart vilket ord som skall användas för översättning, men när detta språk har ett ”finkornigare” ordförråd. I följande exempel så skiljer engelskan på progressiv och avslutad aspekt, vilket franskan (oftast) inte gör i detta fall:

FRANSKA	ENGELSKA
<i>Il s'évanouit</i>	<i>He faints</i> <i>He is fainting</i>

I franska är verbet reflexivt och skiljer inte på de två aspekterna, medan det engelska verbet skiljer mellan två aspektuella former.

5.1.4 Tematisk divergens

En annan skillnad mellan språk är att samma proposition kan skilja sig åt rörande vad vad som är satsens subjekt eller tema.

ENGELSKA	FRANSKA
<i>Jones likes the film</i>	<i>Le film plaît à Jones</i>

I det engelska exemplet är *Jones* tema, medan det i det franska exemplet är *Le film* som utgör satsens tema. Ytterligare ett exempel ges nedan.

ENGELSKA	SPANSKA
<i>I like Mary</i>	<i>María me gusta a mí.</i>

I dessa båda fall skiljer sig språken alltså åt vad gäller temarollen i satserna. Därav termen tematisk divergens.

5.1.5 Kategoriell divergens

Språk skiljer sig också åt såtillvida att samma sak kan uttryckas med olika ordklasser, *kategorier*, i två olika språk.

ENGELSKA	TYSKA
<i>I am hungry</i>	<i>Ich habe Hunger</i>

I engelska uttrycks ”fenomenet hungerkänsla” med ett adjektiv, *hungry*, medan man i tyskan ”har hunger”, ett substantiv.

5.1.6 Head-Switching

Ytterligare en skillnad mellan språk är att olika fraser i normalfallet kan tjäna som satsens huvud. Detta fenomen – s k *head-switching* – kan delas upp i *demotionell* och *promotionell* samt *strukturell divergens*.

5.1.7 Demotionell divergens

Inom relationsgrammatik används termen *demotion* för att beteckna processer där en sats omstruktureras så att en NP flyttas till en 'lägre' position i relationshierarkin. Termen används dock lite allmännare inom andra grammatikmodeller.

ENGELSKA

TYSKA

John likes to swim

Johann schwimmt gern

5.1.8 Promotionell divergens

Termen *promotion* beskriver inom relationsgrammatik motsatsen till demotion (se 5.1.7). Som är fallet med demotion används även termen mindre formellt inom vissa modeller.

ENGELSKA

SPANSKA

John usually goes home

Juan suele ir la casa

5.1.9 Strukturell divergens

Samma sak kan uttryckas med olika syntaktisk struktur i olika språk.

ENGELSKA

SPANSKA

John entered the house

Juan entró en la casa

5.2 Ambiguitet

Ett annat problem är det som rör naturliga språks flertydighet, ambiguitet. Även denna flertydighet kan dyka upp på flera nivåer.

5.2.1 Lexikal ambiguitet

Lexikal flertydighet gränsar givetvis till fenomenet med *wall–Wand/Mauer* ovan, men då detta – och övriga exempel ovan – ändå mest rör finkornigheten hos 'samma' ord, så kan ambiguiteter också uppstå genom det faktum att språk är fulla av *homografer* och *homonymer*, dvs olika ord som råkar stavas och/eller uttalas likadant. Således kan man inte utan vidare översätta det svenska ordet *hov* till engelska, eftersom man inte vet om den korrekta översättningen skulle vara *court* eller *hoof*.

Dessa problem löser man i viss utsträckning genom att domänbegränsa uppgiften. Således arbetar det svensk–engelska tal-till-tal-översättningssystemet *Spoken Language Translation* inom området flygbokning. Detta gör att ett yttrande som

I want to fly to Boston

... undviker den möjliga tolkningen av *fly* som antingen *fluga* eller *gylf*, eftersom man från början begränsat området har man också automatiskt begränsat antalet möjliga tolkningar.

5.2.2 Strukturell ambiguitet

Strukturell ambiguitet råder när man inte vet vilken struktur en viss fras har. Således, för att använda ett slitet lingvistiskt exempel, har meningen:

Han såg mannen på kullen med ett teleskop

... ett stort antal möjliga läsningar rörande vem som har teleskopet, var alla befinner sig och så vidare. Ett sätt att lösa detta vid översättning vore att behålla ambiguiteten och översätta även denna, men eftersom den ambiguitet som ofta uppstår med prepositionsfraser typ *med ett teleskop* inte låter sig överföras till t ex kasusspråk, kan man inte alltid göra på det sättet.

5.2.3 Ordklass

Samma ordform kan tillhöra olika ordklasser. Således tillhör ordet *use* i följande två meningar olika ordklasser:

You must not use abrasive cleaners on the printer casing.

The use of abrasive cleaners on the printer casing is not recommended.

Ett än mer belysande exempel är problemet med hur man skall tolka följande sats:

Flying planes made her duck.

Hur skall detta tolkas? Alternativ i urval är:

<i>Flying planes</i>	<i>made</i>	<i>her</i>	<i>duck</i>
<i>Flygplan</i>	<i>fick</i>	<i>henne</i>	<i>att ducka</i>
<i>Flygande plan</i>	<i>tillverkade</i>	<i>hennes</i>	<i>anka</i>
<i>Att flyga plan</i>		<i>duckning</i>	

Bara i detta fall (fler tolkningar är möjliga!) har vi att göra med $3 \times 2 \times 2 \times 3 = 36$ olika tolkningar. Detta är ytterligare ett exempel på att språktolkning i stor utsträckning har med världskunskap att göra.

5.3 Idiom

Ytterligare ett problem utgör *idiom*, dvs fraser eller uttryck vars mening inte kan utläsas av de i frasen ingående orden, och vars översättning ibland, men inte alltid resulterar i ett liknande uttryck. Jämför följande exempel:

ENGELSKA

SVENSKA

If Sam mends the bucket...

Om Sam lagar hinken...

If Sam kicks the bucket...

Om Sam tar ned skylten...

Inom MT har man funnit att det vanligtvis inte är möjligt att täcka ovanstående och liknade fall med samma regler som övrig översättning sker medelst. Ett sätt att lösa det är att täcka in alla idiom i lexikon och lagra dem som sådana. Problem uppstår då i vissa fall när andra lexikala enheter kan dyka upp inuti idiomerna.

5.4 Sammanfattning

Som framgår av ovanstående smakprov bjuder översättning som område på en hel mängd problem som måste lösas. Som redan nämnts är en stor del av problemen att tillskriva fenomenet världskunskap, och som sådant en smula hänvisat till forskningen kring artificiell intelligens, dvs hur man representerar just denna kunskap. En hel del av problemen är dock rent lingvistiska och borde därför kunna lösas inom MT-domänen. En del av problemen med ambiguiteter kan lösas genom att arbeta inom avgränsade domäner. Märk här att även vi människor har lättare att förstå, alternativt översätta, om vi vet vilket område det rör sig om. En mänsklig översättare skulle inte kunna översätta en mening korrekt om han/hon inte visste om källan var en skämtteckning, en fysiktidskrift eller en veckotidningsartikel.

Således är en strategi inför framtiden just att utarbeta moduler som fungerar bra inom specifika områden, och att man för eventuell generell översättning får välja modul.

6 Olika metoder för översättning: tre huvudstrategier

All översättning går ut på att översätta från *ett* språk, SL (för Source Language) till *ett annat* språk, TL (för Target Language). Det finns givetvis olika metoder och tillvägagångssätt vid maskinöversättning. Den enklaste formen av översättning är givetvis när man översätter från ett språk till *samma* språk, dvs SL = TL. I detta fall går det alldeles utmärkt att bara kopiera texten. **Kopiering** som översättningsmetod har dock den uppenbara bristen att det bara fungerar i detta enda fall, vilket kan ses som en smula begränsat, som översättning betraktat.

I en vackrare värld skulle det gå att använda sig av **ord-för-ord-översättning**, men alla som har lärt sig något främmande språk vet att det inte räcker långt, inte ens mellan så pass näraliggande språk som svenska och norska. Åtminstone krävs något slags språklig analys av de ingående orden för att översättningen skall bli begriplig.

I detta kapitel skall redogöras för olika metoder för översättning.

6.1 Direktöversättning med morfologisk analys: *Transformer Engines*

Direktöversättning är det närmaste man kan komma ord-till-ord-översättning, men orden i detta fall ges en analys som också överförs från SL till TL. Det rör sig alltså inte om uteslutande ett byte av ord i tvenne kolumner. Man talar här om att man förutom själva bytet av lexikalisk enhet (= ord) också begagnar sig av en så kallad **morfologisk transfer**, dvs överföring av den morfologiska information som uttrycks av orden i SL. ”Maskinerna” kallas *transformer engines*.

Schematiskt sker alltså översättning enligt följande modell:



En engelsk SL-text kan således ges följande analys (observera att orden ges i grundform och att den grammatiska formen anges i analysen):

Sats: *Any government is dependent on its supporters*

<i>any</i>	<i>government</i>	<i>be</i>	<i>dependent</i>	<i>on</i>	<i>it</i>	<i>supporter</i>
cat=det	cat=n	cat=v	cat=adj	cat=prep	cat=posspron	cat=n
	num=sg	num=sg			num=sg	num=pl
		pers=3			pers=3	
		tense=pres			sex=neut	

Om man nu skulle översätta detta till franska med bibehållen morfologisk-grammatisk information skulle resultatet (kunna) keda till följande TL-mening:

<i>quelconque</i>	<i>gouvernement</i>	<i>être</i>	<i>dépendant</i>	<i>sur</i>	<i>son</i>	<i>défenseur</i>
cat=det	cat=n	cat=v	cat=adj	cat=prep	cat=posspron	cat=n
	num=sg	num=sg			num=sg	num=pl
		pers=3			pers=3	
		tense=pres			sex=neut	

Detta är emellertid inte grammatiskt korrekt franska. Även mellan så pass närliggande (globalt sett) språk som engelska och franska krävs det minst att en analys på frasnivå sker för omordning av t ex N-ADJ och dylikt.

En fördel med transformermodellen är följande:

- Maskinen kommer inte att bli särdeles brydd av marginella, ovanliga eller konstiga satser i källspråket, eftersom den helt enkelt saknar tillräcklig kunskap för att lägga märke till sådant. Således skulle även meningar som är rent ogrammatiska i SL få en översättning. Detta gör transformermetoden väldigt robust, ett desideratum inom MT.

Nackdelarna är dock fler. Bland annat följande:

- Det är väldigt svårt att garantera att den mening som kommer ut på andra sidan är grammatisk, om ens begriplig. Alltför ofta visar sig helt begripliga meningar i SL få en översättning som är fullkomligt nonsens i TL.
- Transformermodellerna inbegriper så många regler på låg nivå att de är svåra att få grepp om rent vetenskapligt och därför svåra att modifiera eller bygga ut med önskat resultat.
- Metoden är egentligen bara ämnad för enkelriktad översättning mellan två språk och är rent olämplig för flerspråksbehandling.

Direktöversättning som metod har således snabbt visat sig vara otillräcklig, även om en morfologisk-grammatisk analys begagnas.

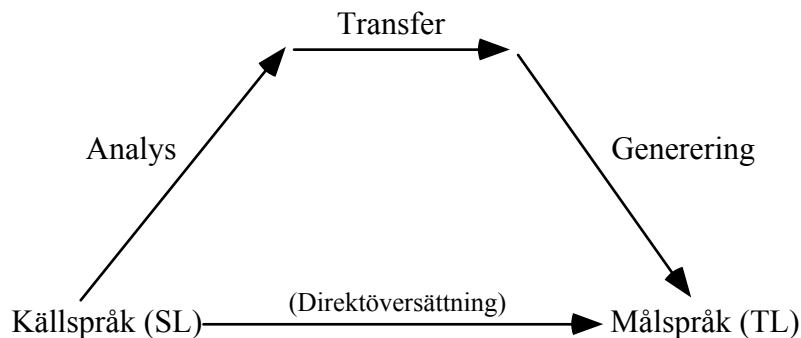
6.2 Översättning via en mellanliggande representation: Transfer

Om nu en direkt översättning SL–TL inte verkar görlig så kan man försöka lösa detta genom att översätta till en representation som är gemensam för de ingående språken, dvs översätta via en mellanliggande representation.

Denna metods arkitektur skulle alltså kunna åskådliggöras således:

SL → MELLANLIGGANDE SPRÅK → TL

Man skulle också kunna åskådliggöra metoden med följande figur:



Som synes översätter man alltså SL till en mellanliggande nivå, från vilken man genererar TL. Vad som krävs i med denna metod är följaktligen tre moduler:

SL → MELLANLIGGANDE SPRÅK → TL

1: Analys 2: Transfer 3: Generering

- Modul 1, **analysmodulen**, analyserar SL till minst konstituentnivå. En förutsättning för detta är (enspråkig) kunskap om SL.
- Modul 2, **transfermodulen**, översätter SL till en mellanliggande representation, som optimalt skall ligga ”mitten mellan” SL och TL, och således förhålla sig neutralt till de bägge språken, dvs, inte vara mer likt vare sig SL eller TL. Transfermodulen innehåller ett slags komparativ grammatik, med kunskap om såväl SL:s som TL:s respektive grammatik.
- Modul 3, **genereringsmodulen**, genererar sedan en mening i TL utifrån den mellanliggande representationen.

Redan här kan nämnas att skillnaden mellan olika förhållningssätt vid MT i själva verket till stor del är en skillnad mellan de olika modulernas respektive storlek. Generellt kan sägas gälla att ju mer SL är analyserat, desto lättare blir transferfasen (översättningen), dvs, desto färre problem kommer man att stöta på.

Denna metod, att gå via en mellanliggande representation som är gemensam för de ingående språken benämnes transferbaserade system och beskrivs i mera detalj i följande stycken.

Som framgått är grunden för denna metod att man analyserar SL ned till en nivå som är gemensam med TL. Således skulle behöva analysera ned satserna i SL till en nivå som gör det klart att ADJ–N i engelska måste vändas till N–ADJ i franska. Däremot så skulle man inte behöva vända på SVO¹ mellan franska och engelska (om inte O är ett pronomen i franska), eftersom engelska och franska har samma grundläggande ordföljd.

När man nått fram till denna mellanliggande nivå så översätts/-förs den till TL genom transfer. Denna metod kräver att viss funktionell analys av SL har gjorts.

Transfer i detta fall involverar två olika fenomen, dels *lexikal transfer*, dvs överföring av lexikala enheter, dels *strukturell transfer*, dvs överföring till den struktur i TL som bäst motsvarar den i SL analyserade strukturen.

6.2.1 Lexikal transfer

Lexikala transfer krävs för att byta ut SL-orden till motsvarande TL-ord, ett ganska självklart steg i all översättning. Ett uppenbart problem förknippat med detta är att olika språk, som tidigare nämnts, inte har helt och hållet konguenta ordförråd och inte via dessa kategoriserar världen på samma sätt. Den finkornighet med vilken ett språk beskriver världen på ordnivå motsvaras ofta inte av finkornigheten i ett annat språk. Denna inkongruens kan dock yttra sig på olika sätt.

I de fall det faktiskt råder ett **ett-till-ett-förhållande** mellan SL och TL har vi inga problem. Orden betyder i detta fall samma sak, och det finns lika många ord – dvs ett – i såväl SL som TL för att beskriva fallet.

De fall då det råder ett **många-till-ett-förhållande** mellan SL och TL har vi heller inga problem. Granulariteten i SL är högre än i TL, vilket gör att flera givna ord i SL översätts till samma ord i TL. Ett exempel här är t ex då man skall översätta *Wand* eller *Mauer* i tyska till engelska, då bägge orden skulle resultera i det engelska ordet *wall*.

Då det råder ett **en-till-många-förhållande** mellan SL och TL stöter man emellertid på problem. Ett exempel här är då man skall översätta engelska *wall* till tyska. Hur skall man veta om det rör sig om *Wand* eller *Mauer* i det aktuella fallet?

Ordval sker i vissa fall också **kontextuellt**. Det engelska *library* kan ge tyska *Bibliothek* om det är akademiskt bibliotek men *Bücherei* om det är ett allmänt bibliotek. På samma sätt kan det engelska *grandmother* alludera till en *farmor* eller en *mormor*. Båda dessa exempel visar på de fall då det ena språket har en hyperonym term, medan det andra i stället har två hyponyma begrepp, men saknar det hyperonyma.

6.2.2 Strukturell transfer

Ett annat slags transfer är den så kallade *strukturella transfer*. Den behövs när SL:s syntaktiska struktur inte motsvaras direkt av TL:s dito. Generellt gäller att ju djupare analysen av SL är, desto mer sällan skall detta problem uppstå – åtminstone i teorin – eftersom analysen ju har som syfte att neutralisera just skillnaderna mellan SL och TL.

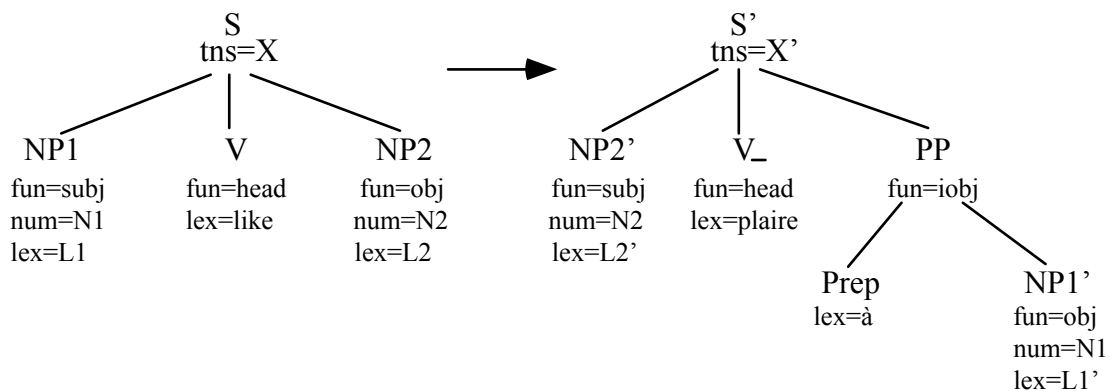
¹ Man brukar klassificera språk efter vilken ordning subjekt (S), objekt (O) och predikatsverb (V) har i deklarativa satser.

Vissa strukturella problem är lättare att lösa än andra. Ett ganska lättlost problem belyses i följande exempel:

Jones likes the film → *Der Film gefällt dem Jones*
Le film plaît à Jones

Det engelska ordet *like* motsvaras bra av det tyska *gefallen* och det franska *plaire*, men de olika språkens respektive strukturer matchar inte. Subjektet till *like* måste göras till indirekt objekt i franska respektive dativ i tyska.

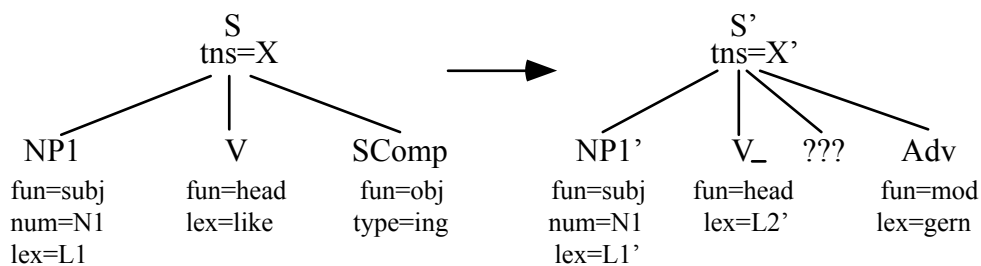
Följande figur belyser det franska exemplet:



Ett något svårare problem utgörs av följande exempel:

John likes to swim → *John schwimmt gern*

I detta fall har vi att göra med head switching. Följande figur åskådliggör detta:



Om detta dessutom innehåller adverbial så måste dessa ”skickas med” på vägen upp till toppnivån.

John likes to swim with his friends in the summer

↓

John schwimmt im Sommer mit seinem Freunden gern

Transfermodulen blir i detta fall inte generell i någon vidare mening, utan återspeglar SL:s och TL:s struktur.

6.2.3 Sammanfattning

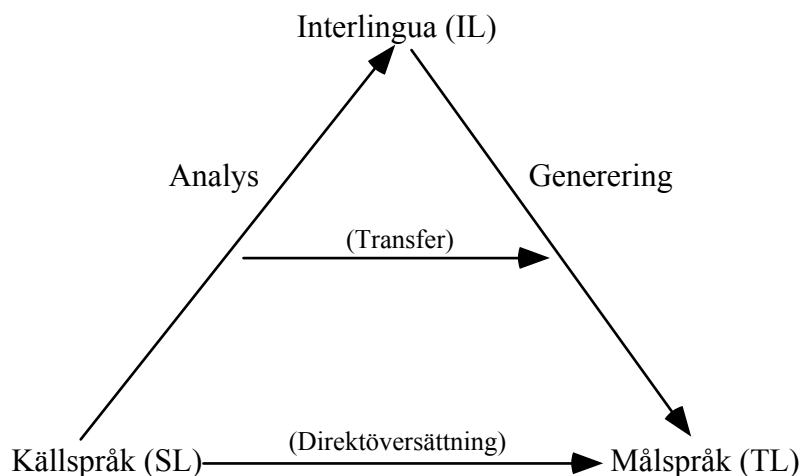
Transfer som metod har både för- och nackdelar. Till fördelarna hör att man inte behöver analysera djupare än till den nivå där SL och TL är 'lika'. Detta gör de enspråkiga modulerna mindre, dvs mappningen $SL \rightarrow IL$ (dvs, interlingua) respektive $IL \rightarrow TL$. Till nackdelarna hör att man för varje nytt språk som inkluderas måste lägga till en ny enspråkig modul (i bägge riktningar!). Detta leder till en att man får en generell tillväxt på $N \times N - 1$.

Dessutom måste man vid tillägg av ett nytt språk antagligen justera IL-modulen, eftersom det nya språket säkerligen uppvisar drag som de tidigare ingående språken inte haft. Detta har som resultat att ett transfersystem ofta får låg flexibilitet.

6.4 Interlingua-system

Som vi såg kräver transfermetoden att man för varje nytt språk som inkluderas skapar två nya moduler, en för mappningen $SL \rightarrow IL$, samt en för mappningen $IL \rightarrow TL$. Man tvingas alltså mer eller mindre göra om hela transfermodulen för att anpassa den till det nya språket. Detta leder till att inkluderandet av nya språk medför ett exponentiellt växande system.

I "rena" interlinguasystem så eftersträvar man ett IL som skall vara "universellt". Således skulle följande figur representera filosofin bakom ett rent interlingua-system:



Denna figur – som ibland går under namnet *översättnings-A:t* – sammanfattar i princip strategierna bakom översättning. Ju högre upp i pyramiden man analyserar, ju mer språkoberoende blir den mellanliggande representationen, och toppen på pyramiden skall utgöra en helt språkoberoende representation, det så kallade *interlinguat*.

I ett interlingua-system översätts SL till en representation som är identisk med den representation från vilken generering till TL skall ske, detta till skillnad från transfersystem, där SL översätts till en representation som 'stuvats om' på ett eller annat sätt innan generering äger rum. Fördelen med ett interlingua med denna – önskvärt optimala – universalitet, är att varje nytt språk som läggs till systemet bara kräver anpassning till IL, medan IL självt inte behöver anpassas eller ändras. Att lägga till nya språk till systemet bara medför ett linjärt växande system.

Det finns dock nackdelar med denna ideologi. En första nackdel är att analys och generering måste hållas isär, dvs man kan inte analysera med ett speciellt TL i sinnet, och man kan inte titta tillbaka till SL vid generering. Ett större problem är att IL måste innehålla unionen av alla de drag – grammatiska, lexikala mm – som mänskliga språk uppvisar. Att konstruera ett sådant språk låter sig inte göras i en handvändning.

Förslagen till vad som skulle kunna tjäna som IL i ett interlinguasystem har varit flera, och jag skall i det följande beskriva några av dessa.

6.4.1 Olika förslag till IL

Ett förslag har varit att använda sig av **sats-** eller **predikatlogik**. Logik utvecklades ju i mångt och mycket för att komma ifrån mänskliga (naturliga) språks begränsningar, och torde ju vid en första anblick verka lämpligt som ett ”rent” IL. Man kan visserligen redan här påpeka att predikatlogik är VSO, varför det åtminstone i denna bemärkelse är mer likt t ex arabiska än germanska språk.

Nåväl, hur skulle då en en predikatlogisk IL-representation se ut? En mening som

Any government is dependent on its supporters

... skulle kunna se ut på följande sätt:

all(X), goverment(X), indefinite(Y), plural(Y), support(Y,X,T),
depend-on(X,Y,T), timeless(T).

Erfarenheten har visat ett det är lätt att nå en slik representation vid analys. Däremot har det visat sig mer än svårt att från en sådan representation generera en text i något naturligt språk.

Ett annat förslag som förts fram har varit **Chomskys djupstrukturer**. Dessa ansågs inom tidig MT vara lovande, men anses nu inte vara tillräckligt abstrakta, eftersom de i alltför hög grad är orienterade mot TL:s ytstrukturer.

Ytterligare ett förslag är Fillmores **Case Grammar** (1968). Besläktad med predikatlogisk representation skulle meningen

Sam killed Kim

... kunna ges följande representation:

KILL(Sam, Kim)

Eftersom man inte kan räkna med att alla språk har ett transitivt verb *kill*, skulle en mer ”detaljerad” eller djupare analys av meningen ovan kunna plocka isär själva verbet *kill* i dess beståndsdelar:

CAUSE(Sam(BECOME(NOT(ALIVE, Kim))))

Som bekant har språk heller inte samma tempus-system, varför öven själva tidsaspekten måste lyftas fram i analysen. Detta skulle kunna ske på följande sätt:

$$\exists e_1, e_2 [\text{SUBJ}(e_1, \text{Sam}) \ \& \ \text{CAUSE}(e_1, e_2) \ \& \ \text{DYING}(e_2) \ \& \ \text{SUBJECT}(e_2, \text{Kim})]$$

Den grammatiska analysen ovan skulle kunna fördjupas ytterligare, och en representation som den som följer har visat sig motiverad:

$$\begin{aligned} &\exists e_1, e_2 [\text{SUBJ}(e_1, \text{Sam}) \ \& \ \text{CAUSE}(e_1, e_2) \ \& \ \exists s_1 s_2 [\text{COMING-ABOUT-OF}(e_2, s_1) \ \& \ s_1 \\ &= \neg \text{HOLDS}(s_2) \ \& \ \text{BEING-ALIVE}(s_2) \ \& \ \text{OBJ}(s_2, \text{Kim})]] \end{aligned}$$

Detta senare exempel skulle kunna utläsas som:

Det finns en händelse e1 vars agent är Sam, och denna händelse orsakar en händelse e2 som handlar om att ett tillstånd s1 skapas. Detta tillstånd s1 är ekvivalent med att tillstånd s2 - att Kim lever - inte gäller.

Att använda sig av Case Grammar verkade till en början attraktivt. Man kunde explicit redogöra för innehåll i satser, och representationerna var sådana att man faktiskt kunde överföra dem till naturliga språk. Emellertid finns det även här problem av oöverstiglig natur. För det första har ingen kunnat enas om vilka semantiska predikat som skall användas, dvs om man skall använda sig av:

KILL

... vilket inte räcker, eller:

CAUSE
BECOME
NOT
ALIVE

... eller återigen:

SUBJ
CAUSE
COMING-ABOUT-OF
HOLD
BEING-ALIVE

... etc. faktum är att man har visat att de semantiska predikatet för ett ord som t ex *duva* måste innehålla ett predikat [+DUVA] eller [+DUVLIK] eller dylikt. Således resulterar alla försök till att definera semantiska predikat till att man förr eller senare ”biter sig i svansen” (Jfr Fodor 1977). Ytterligare problem med interlinguasystem är att man inte bara måste ta hänsyn till ordföljd (som för språk som engelska och svenska) utan också syntaktiska funktioner, kasus, tema-remå, bestämning etc. Skillnaden mellan ord som *like* och *gern* i tidigare exempel kan relativt lätt beskrivas i ett transfersystem med en strukturell transfer-regel med en mappning mellan SL och TL. Ett IL kan antingen välja *like* som huvud eller *gern*, men ingetdera valet vore språkneutralt. Detta leder givetvis till frågan om vad som är språkneutralt. En lösning här har varit att man ofta väljer det som är vanligast i majoriteten av språk. Problemet med att finna en neutral representation har fått olika lösningar, varav några presenteras i följande avsnitt.

6.4.2 Abstraktion respektive normalisering

I frågan om hur man skall lösa problemet med att finna en neutral representation för t ex ordföljd brukar man skilja på så kallad *normalisering* och så kallad *abstraktion*.

Vid **normalisering** antar man att det finns *en* ordning, och denna tillämpas sedan på alla språk. Som exempel skulle man kunna sätta ordföljden till att vara VSO. Ett annat exempel skulle kunna vara att man i systemet antar att alla NP:n föregås av artiklar, vilka inte realiseras i alla språk. Fördelar med denna metod är att man undviker språkpar-beroenden. Alla språk matchas mot en och samma struktur, även om avståndet till denna varierar. En nackdel är att den lingvistiska motiveringen kan bli en smula ”bisarr” om man vid översättning mellan engelska och franska (SVO-språk) måste arbeta med en VSO-struktur då bägge språken är SVO, eller om man vid översättning mellan latin och ryska måste arbeta med artiklar, som bägge språken saknar.

Den andra nämnda metoden är **abstraktion**. Man antar här att det finns en nivå som är *oordnad*. Detta är för övrigt något som tillämpas inom t ex LFG och dependensgrammatik. En fördel med detta är att på detta sätt undviker konstiga motiveringar, som ovannämnda exempel. Inget språk kan t ex få en konstig ordföljd på denna mellanliggande nivå eftersom man helt abstraherat iväg från ordföljd. En stor nackdel här är dock att *all* information – som t ex ordföljd – som finns i SL måste omkodas och ”tas med” i alla fall, på något annat sätt.

Skillnaden mellan de bägge metoderna är att den ena, abstraktion, uppnår transfer genom en universell närvaro av artiklar, medan den andra, normalisering, undviker transfer genom att kräva en universell frånvaro av artiklar. Att välja mellan de bägge synsätten blir mycket en fråga om vad man finner mest ”onaturligt”, översättning mellan t ex ryska och latin genom skapande av artiklar, eller översättning mellan engelska och svenska med en nivå utan artiklar.

6.4.3 Problem med interlinguasystem

Sammanfattningsvis har interlinguasystem framför allt två problem att brottas med:

- Att välja den mest lämpliga, språkneutrala, representationen.
- Att komma fram till hur man skall få fram – och ta med sig – all nödvändig information ur SL.

Medan ett transfersystem inte har några problem med en-till-en-mappningar av ord, får ett interlinguasystem problem om bara *ett* av de ingående språken har fler än en form för ett visst ord. Varje distinktion som kan göras i *ett* språk måste representeras explicit i IL. Detta kräver givetvis ofantligt mycket arbete, även för näraliggande språk. Om spanska är ett av de ingående språken så måste man skilja på ben på människor, djur, möbler (etc) även om man just håller på att översätta mellan engelska och tyska. Om japanska är med så måste man hålla isär åtta olika *wear* (japanska distinktioner). Dessutom så skiljer engelska på att *bära* och att *ta på sig* kläder, vilket japanska inte gör, varför IL måste innehålla 16 olika representationer bara för ordet *wear* om engelska och japanska är inkluderade.

På grund av dessa (och andra) problem interlinguasystem försöker man heller inte uppnå ett rent IL som beskrivet ovan, utan man använder sig i stället av något som mer liknar transfersystemens mellanliggande representationer. Generellt kan nämnas att existerande interlinguasystem ofta i hög grad förlitar sig på kontextuell kunskap!

6.4.5 Interlingua med redan existerande språk

Det finns interlinguasystem som använder sig av ett existerande språk som IL. Kraven på ett sådant IL är bland andra att det måste vara:

- Oambiguöst.
- Konsistent.
- Regelbundet (ordförråd, syntax mm).
- Inga homografer eller polysemer.
- Kunna innefatta flertalet konceptuella skillnader i mänskliga språk.

Dessa krav uppfylls inte av något naturligt språk, men vissa språk som skapats för internationell kommunikation betraktas ibland som möjliga kandidater. Man har bland annat försökt använda en variant av esperanto som IL. Det är dock tveksamt om ett sådant system kan kallas för ett interlinguasystem, eftersom det är svårt att se på vilket väsentligt sätt en översättning engelska–esperanto skulle skilja sig från en översättning engelska–franska!

6.4.6 Restrigerade interlinguasystem

Även om det kanske är omöjligt (eller åtminstone inte görligt) med ett riktigt, ”rent” interlinguasystem, så kan man konstruera system som är interlinguala för de språk som ingår i systemet. Sådana system skulle använda sig av ett IL som var helt neutralt i förhållande till de ingående språken, men samtidigt kunna dra nytta av deras likheter. T ex skulle man kunna föreställa sig ett romanskt system. Inom EUROTRA-projektet har upptäckts att det har varit möjligt att etablera semantiska ”euroversaler”, dvs interlinguala lexikala enheter för EU:s språk.

6.5 Sammanfattning av olika strategier

Vi har ovan belyst lite av den grundläggande problematiken som rör översättnings-A:t, och olika sätt att förhålla sig dessa problem samt olika förslag till lösningar. I tabellen nedan sammanfattas för- respektive nackdelarna för direkt-, transfer- respektive interlingua-baserad översättning.

	Direkt	Transfer	Interlingua
PRO	Krävs väldigt lite teoretisk kunskap/analys. Robusta. Allt får översättning.	Anses vara den görligaste vägen att komma vidare. Flexibel metod vad gäller användning av olika hjälp-metoder. Analysmodulerna mindre komplexa än hos IL-system. Fokus ligger på kontrastiv lingvistisk kunskap.	Lätt att lägga till nya språk. (För varje nytt språk krävs bara två nya moduler.) Färre moduler än transfer. Störs inte av språk-specifika undantag. Om IL verkligen är oberoende så kommer översättningar bli mycket idiomatiska.
CON	Stor belastning på de enspråkiga lexika. <i>Ad hoc</i> -behandling av strukturella skillnader. Översättningar av låg kvalitet.	Svårt att lägga till nya språk. Fler moduler än IL.	Ett sant IL ligger bortom lingvistisk teori och kunskap. All kontrastiv kunskap ligger dold i de monolinguala analys- och genereringsmodulerna, den bilinguala transfern och TL-genereringen. Kräver unionen av alla språk (T ex ordförråd!) Antagligen kan en SL-text ge flera IL-representationer, vilka var och en kan ge flera TL-texter. Om IL tillräckligt oberoende för avlägsna språk så kommer det att vara onödigt komplext för nära-liggande språk.

Sammanfattningvis kan sägas att de flesta existerande system inte är rena transfer- eller interlinguasystem, utan kombinerar strategier från bägge angripssätten.

7 Alternativa metoder

De översättningmetoder som beskrivits ovan är inte de enda som förekommer, utan alternativa förhållningssätt har också sett dagens ljus, och jag skall i detta stycke kort presentera några av dessa.

7.1 Exempelbaserad översättning

Som komplement till den traditionella analys-transfer-genereringsmodellen har föreslagits exempelbaserad översättning, bland annat föreslaget av Nagao (1984). Enligt denna metod skulle man via tvåspråkiga korpora (manuellt översatta) erhålla kunskap om vad som översätts som vad i olika kontexter. Huruvida ett ord som *field* (i engelska) skulle översättas till *champ*, *domaine* eller *bassin* (i franska) avgörs av

frasfrekvensen av de mest lika fraserna i korpusen. 'Likhet' i detta fall kräver distansmätning, vilket för närvarande oftast sker genom klassificering i semantiska hierarkier (typ hyperonymer osv).

Ett exempelbaserat system för talspråk kallat **TDMT**, för *Transfer-Driven Machine Translation* presenteras av Furuse och Iida (1992). Ramverket bygger på flernivå-kunskap. Andra liknande metoder som föreslagits är bl a *Example-Based MT* (Sumita & Iida 1991), *Memory-Based MT* (Sato & Nagao 1990) och *Analogy-Based MT* (Sadler 1989).

Observera att exempelbaserad MT inte är detsamma som korpusbaserad MT. Korpusbaserad MT innebär att man utvecklar system för översättning av texter i en viss korpus och sedan används för översättning av andra texter i samma korpus.

Exempelbaserad MT kan integreras i vilken som helst av de tidigare nämnda systemen: direkt, transfer eller interlingua.

Exempelbaserad översättning har vissa klara fördelar. TMDT är ett system för översättning av talspråk. Talspråk kännetecknas av att det avviker mycket från konventionell grammatisk beskrivning varför ett exempelbaserat system är mycket lämpligt i detta fall. Exempelen valideras genom statistisk evidens. Man använder sig således av en *best-matching mechanism* i stället för en *exact matching mechanism*. Exempelbaserad MT är också mycket robust. Om systemet gör fel någonstans kan detta lätt korrigeras genom att lägga till ett nytt exempel. Metoden är också snabb eftersom distansmätningen baseras på en enkel algoritm. Man kan kontrollera och balansera vikterna lätt. Ytterligare en fördel är som sagt att man utan vidare kan kombinera exempelbaserad MT med någon av de andra metoderna. Det har visat sig att exempelbaserade metoder är synnerligen lämpliga för översättning av komplexa NP:n.

Det finns emellertid även nackdelar med exempelbaserad översättning. Det har visat sig dyrt att bygga en stor exempelbaserad korpus. Man kan också invända mot metoden att den inte baserar sig på lingvistisk kunskap, utan enbart på statistisk evidens.

7.2 Shake-and-bake-system

En annan metod som föreslagits är det så kallade *shake-and-bake*-metoden, tex av Whitelock (1992).

Shake-and-bake utgör ett filter- (*constraint*) baserat synsätt på språk. Lingvistisk struktur bestäms genom ackumulering av beskrivningar eller begränsningar. Detta leder till deklarativa och kompakta grammatiker. Man kombinerar inom shake-and-bake-översättningen interlingua-representationer med bibehållna lågnivåbeskrivningar. En föregångare till detta synsätt var Landsbergen (1987) som hävdade att den tvåspråkiga kunskap som krävdes för att beräkna ekvivalenta uttryck vid översättning skulle ges som korrespondenser mellan grammatiker snarare än mellan språk.

Ett karaktärsdrag inom shake-and-bake är att man drastiskt underspecificerar information om TL-struktur (enl Sadler), SL-text (enl. Whitelock). Man arbetar med översättnings-ekvivalenser vilka definieras som "säckar" (*bags*) av lexikala enheter. SL-analysen bestämmer/fastslår vilka säckarna på ena sidan analysen är, och shake-and-bake-generering kombinerar fram den ekvivalenta säcken på andra sidan, dvs TL.

Genereringen sker i detta fall helt fritt enligt TL:s grammatik. Lexikala enheter ses som *tecken*. Detta innebär att de klassificerar ett uttryck simultant i flera dimensioner, som ortografiska, syntaktiska, semantiska. Den grammatik som används är således helt skapad utifrån templat som introduceras av lexikala enheter. Detsamma gäller de logiska formerna (LF), vilka deriveras uteslutande från lexikala enheter. Man representerar sats-betydelse som en kombination av ordbetydelser och ett antal filter på de erhållna LF-variablerna.

Det enda som betraktas som betydelsebärande är de lexikala elementen, och tvåspråkig kunskap och bilinguala korrespondenser definieras uteslutande mellan lexikala tecken deras respektive associerade betydelser. Ekvivalens mellan två sådana lexikala enheter behöver inte vara ett-till-ett, som följande exempel belyser:

$$\begin{aligned} \{pay, attention, to\} &= \{faire, attention, à\} \\ \{as, as\} &= \{aussi, que\} \end{aligned}$$

Ekvivalens uttrycks följaktligen som som ekvivalens mellan säckar av lexikala enheter. En mening representeras av en säck med lexikala enheter som kallas 'bas' (*base*). Två baser ses som ekvivalenta om de utgör unionen av ekvivalenta säckar. Två uttryck är översättningsekvivalenta om de har härledningars vars respektive baser är ekvivalenta säckar och de inkluderar samma begränsningar vad gäller LF-variablerna.

Shake-and-bake kan betraktas som en modifiering av en *shift-reduce parser* för binära grammatiker. I en shift-reduce-parser plockas upprepat ett element av en indatasträng, slås upp i ett lexikon varefter resultatet slängs på en stack. De två översta elementen på stacken kan sedan kombineras ihop enligt grammatiken, varefter resultatet återigen slängs på toppen av stacken och proceduren anropas rekursivt. En skillnad mellan shake-and-bake och shift-reduce är att i shake-and-bake så spelar ordningen på elementen ingen roll.

Ett exempel på hur ett översättningsekvivalens mellan det engelska ordet *love* och det franska ordet *aimer* ser ut är följande:

$$X_E + X_F$$

$$\langle X_E \text{ cite} \rangle = \text{love}_v$$

$$\langle X_F \text{ cite} \rangle = \text{aimer}$$

$$\langle X_E \text{ sem index} \rangle = \langle X_F \text{ sem index} \rangle$$

$$\langle X_E \text{ sem exp index} \rangle = \langle X_F \text{ sem exp index} \rangle$$

$$\langle X_E \text{ sem obj index} \rangle = \langle X_F \text{ sem obj index} \rangle$$

I detta fall råder det ett ett-till-ett-förhållande mellan de lexikala enheterna. Metoden låter sig dock lätt anpassas till andra förhållanden. Vi betraktar ett exempel som inkluderar argument-byte (*argument switching*):

Mary likes Frances

Françoise plaît à Marie

I detta fall subkategoriserar *like* och *plaire* annorlunda, och subjekts- respektive objektsfunktionerna är omkastade. Ekvivalensen för dessa enheter ser ut som följer:

$$X_E + X_F \& X_F'$$

$$\langle X_E \text{ cite} \rangle = \text{like_v}$$

$$\langle X_F \text{ cite} \rangle = \text{plaire}$$

$$\langle X_F' \text{ cite} \rangle = \grave{a}_1$$

$$\langle X_E \text{ sem index} \rangle = \langle X_F \text{ sem index} \rangle$$

$$\langle X_E \text{ sem exp index} \rangle = \langle X_F \text{ sem exp index} \rangle$$

$$\langle X_E \text{ sem obj index} \rangle = \langle X_F \text{ sem obj index} \rangle$$

$$\langle X_E \text{ sem exp index} \rangle = \langle X_F' \text{ sem index} \rangle$$

Här antas att uttrycket $\langle X_F \text{ sem obj} \rangle$ plockar ut ordet *like*:s semantiska objekt, vilket motsvaras av ordet *plaire*:s semantiska subjekt. Prepositionen *à* i detta fall inkluderas helt enkelt genom att det anges att ett ord på engelska motsvaras av två ord på franska.

Argumentbyten som ovan utgör inget problem för gängse transfersystem. Däremot så är ofta *head switching* problematiskt för transfersystem. Head switching är inte mer problematiskt än huvudbyte för shake-and-bake-metoden. Således skulle de bägge satserna

Jan zwemt graag

John enjoys swimming

... helt enkelt beskrivas som ekvivalenser mellan simplexa uttryck, sålunda:

$$X_E \& X_E' + X_N$$

$$\langle X_E \text{ cite} \rangle = \text{enjoy}$$

$$\langle X_E' \text{ cite} \rangle = \text{prespart}$$

$$\langle X_N \text{ cite} \rangle = \text{graag}$$

$$\langle X_F' \text{ cite} \rangle = \grave{a}_1$$

$$\langle X_E \text{ sem index} \rangle = \langle X_N \text{ sem index} \rangle$$

$$\langle X_E \text{ sem exp index} \rangle = \langle X_N \text{ sem exp index} \rangle$$

$$\langle X_E \text{ sem obj index} \rangle = \langle X_N \text{ sem obj index} \rangle$$

$$\langle X_E' \text{ sem index} \rangle = \langle X_N \text{ sem index} \rangle$$

De fulla säckarna får följande utseende:

$$\{\text{jan,pres,zwemen,graag}\} + \{\text{john,pres,enjoy,prespart,swim}\}$$

I och med att en tvåspråkig lexikograf inte behöver specificera ekvivalenser mellan abstrakta strukturer på en enda representationsnivå (som i transfersystem), utan i stället fastslår ekvivalens mellan säckar med multidimensionella tecken, anser Whitelock att shake-and-bake erbjuder ett steg mot integrering av traditionellt (rationalistiskt) och modern, statistiskt (empiricistiskt) förhållningssätt visavi MT. Det har visats att shake-and-bake är ett NP-komplett problem (Brew 1992), varför man inte kan erhålla någon generell algoritm.

7.3 Abduktionsöversättning

Ytterligare en metod för översättning är den så kallade abduktionsmetoden, som presenteras i Hobbs & Kameyama (1990). Avstampen för metoden är att språk dels uttrycker "samma" information på väldigt olika sätt, dels skiljer sig åt, ofta drastiskt, med avseende på vilka lingvistiska distinktioner som måste göras. Detta leder till att antagandet att all information som måste ges i TL står att finna i SL faller. Ett exempel som ges är att vid översättning från engelska till japanska kan man ingenstans i engelskan hitta information om de artighetsformer som måste användas i den japanska översättningen.

För att kunna göra adekvata översättningar krävs en gedigen förståelse av SL-texten. Maskinöversättning kräver således att de lämpliga inferenserna dras och att SL-texten interpreteras, i djupare mening.

Med abduktion avses "inferens till den bästa förklaringen". Skillnaden mellan deduktion, induktion och abduktion är som följer:

Deduktion

Från ett faktum $p(A)$ och en generell regel $(\forall x)p(x) \supset q(x)$ konkluderar vi $q(A)$.

Induktion

Från ett antal instanser $p(A)$ och $q(A)$ (plus ev. annat) konkluderar vi $(\forall x)p(x) \supset q(x)$.

Abduktion

Från $q(A)$ och $(\forall x)p(x) \supset q(x)$ konkluderar vi $p(A)$.

Man menar att mycket av vårt sätt att förstå världen baseras på abduktion. När det som skall analyseras är en text måste vi tillhandahålla den bästa inferensen för att texten skall vara sann. Det ramverk som presenteras kräver integrerad behandling av syntax, semantik och pragmatik. Det 'abduktiva ramverket' presenteras sålunda:

För att tolka en mening framskrider man på följande sätt:

Prove the logical form of the sentence, together with the constraints that predicates impose on their arguments, allowing for coercions, merging redundancies where possible, making assumptions where necessary.

Ett yttrande, menar man, *stands with one foot in mutual beliefs and one foot in the speaker's private beliefs*. Vad som krävs är kunskap om diskursen och situationen. En översättning bör i idealfallet leda lyssnaren till samma situation som meningens tolkning. För att översätta använder man sig av axiom som specificerar transferrelationer. Dessa axiom inkluderar världskunskap, lexikala ambiguiteter (lexikal disambiguering uppstår som en biprodukt av tolkningen av ett yttrande), direkta relationer mellan språken samt relationer mellan de bägge språken och djupare, interlinguala representationer. Således inkluderar systemet olika nivåer i beskrivningen, från yttnivåer till djupa relationer på interlingua-nivå. Detta leder till att man kan välja graden på den lingvistiska abstraktionsnivån, och skapa transferregler på lämplig nivå.

Hobbs och Kameyama identifierar fyra områden där översättning mellan engelska och japanska är särskilt svår:

- Lexikala skillnader.
- Artighetsformer, *honorifica*.
- Bestämmdhet och numerus.
- Kontextkänslig 'informationsstruktur'.

Man menar att abduktionsramverket borde klara av de första tre punkterna. (Vad som egentligen avses med punkt 4 förklaras inte i närmare detalj!).

Lexikala skillnader kan helt enkelt specificeras i axiomen. (Observera att japanska inte skiljer på föremålet/egenskapen!)

$$(\forall x)\text{water}(x) \wedge \text{cool/cold}(x) \text{ mizu}(x)$$

Artighetsformer uttrycks mycket olika i engelska och japanska. I japanska är artighet oftast grammatikaliserat och/eller lexikaliserat. Vid översättning av t ex datormanualer stöter man inte på detta problem, men om det är affärsbrev som skall översättas blir problemet mycket aktuellt. Hobbs och Kameyama menar att även *honorifica* kan kodas axiomatiskt när information om t ex hierarki framgår av texten. Man skulle sålunda kunna inkludera predikat som HÖGRE-STATUS i axiomen. På detta vis skulle man lätt kunna begränsa bruket/användningen av t ex vissa verb.

Bestämmdhet och **numerus** ställer till problem mellan engelska och japanska. Det japanska ordet *ki* kan översättas med *a tree*, *the tree*, *trees* eller *the trees* på engelska. Här menar man att lösningen ligger i en tre-stegsmodell, där steg ett letar efter referenten till den japanska NP:n, steg två kontrollerar om den japanska NP:n refererar till en mängd av två eller fler, och steg tre kontrollerar om enheten är "ömsesidigt identifierbar", i vilket fall den översätts i bestämd species. Med "ömsesidigt identifierbar" förstås att beskrivningen som den japanska NP:n tillhandahåller i ömsesidigt känd, och att det existerar en enskild mest framträdande sådan enhet.

Hobbs och Kameyama avslutar med att påpeka att vad som krävs för maskinöversättning är dels djuptolkning när sådan krävs, översättning från SL til TL på olika, lämpliga, nivåer, samt tillgång till världskunskap, kontext och talaktssituation. Abduktions-översättning klarar av allt detta, menar man. Ett system som använder abduktion är **TACITUS** vid SRI i Menlo Park. Det syftar till att förena syntaktisk, semantisk och pragmatisk kunskap.

7.4 Kunskapsbaserad översättning

Lingvistiska skillnader och språkförståelse baseras i hög grad på kulturella skillnader. Det har därför föreslagits att översättning borde baseras på icke-lingvistiska, konceptuella representationer, eller på *betydelser* baserade på *förståelse* av text. Den så kallade kunskapsbaserade metoden att lösa detta baserar sig på att man skapar en domänmodell inom vilken man har/kan specificera inte bara rent lingvistisk kunskap, utan även detaljerad semantisk och pragmatisk kunskap.

Om man antar att betydelse och förståelse är universella skulle en slik representation utgöra ett sant IL. Ett sådant system kräver per definition världskunskap, och de flesta samtida system inriktar sig inte på detta slags representationer. Exempel är dock **KBMT** och **KANT** vid Carnegie Mellon University.

7.5 Strukturkorrespondensöversättning

Översättning genom utnyttjande av strukturella likheter på olika lingvistiska nivåer föreslås av Kaplan et al (1989). De benämner sin metod *co-description*, vilket skall belysas som följer.

Metoden går i princip ut på att relationen mellan SL och TL definieras som en mängd korrespondensrelationer, i stället för att tillhandahålla härledningar eller procedur tekniker för konverteringen från SL till TL. En fördel med detta är att mappningen mellan SL och TL blir avhängig information från flera olika nivåer i den lingvistiska beskrivningen, samtidigt som man behåller modulariteten hos de lingvistiska komponenterna och håller isär SL:s och TL:s respektive grammatiker och lexika.

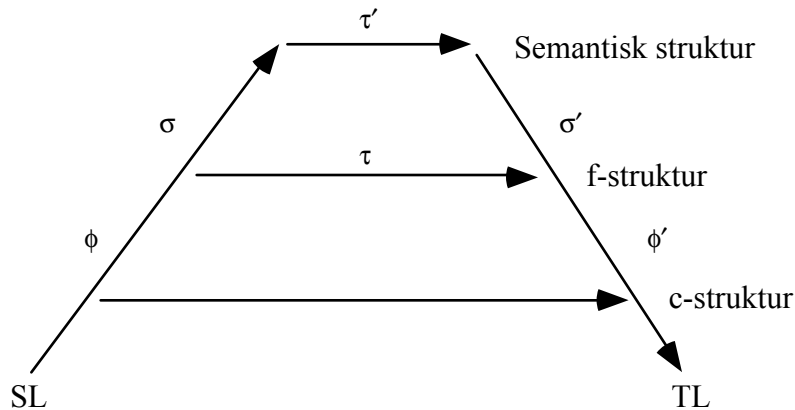
Nyckelord inom metoden är *struktur*, *strukturell beskrivning* och *strukturell korrespondens*.

Kaplan et al menar att såväl transfer- som interlinguasystem teoretiskt har möjlighet att lyckas bra om de inkluderar den lingvistiska kunskap/information som krävs i det aktuella systemet. De lingvistiska nivåer som man måste inkludera information om är oftast de traditionella: fonologi, morfologi, syntax, semantik, diskurs etc. Kaplan menar emellertid att ett problem är att inget enskilt av dessa områden tillhandahåller tillräcklig information för att möjliggöra adekvat information. Som exempel påpekar man att medan *Government & Bindings* D-strukturer innehåller information rörande predikatargument-relationerna i en sats så ger de ingen som helst information om konstituentordningen på ytan, vilket behövs för att hålla reda på tema-remå i en sats. Den metod Kaplan et al föreslår tillåter dem att fastslå simultana korrespondenser mellan flera nivåer av källa-mål-representationer.

Metoden man använder baseras på de egalitets- och beskrivningsbaserade mekanismerna inom LFG, *Lexical-Functional Grammar* (Kaplan & Bresnan 1982). Inom LFG tilldelas varje mening två nivåer av syntaktisk representation, en konstituentstruktur, *c-struktur*, och en funktionell struktur, *f-struktur*. C-strukturen är ett traditionellt frasstrukturträd (i princip kontextfritt), medan f-strukturen är en finit hierarkisk funktion, givet som en mängd booleska funktion-argument egaliteter. De bägge strukturerna belyser olika bitar av meningens information, såtillvida att c-strukturen visar ordningen på ord och fraser, medan f-strukturen belyser funktioner som subjekt, objekt osv.

Inom LFG används en funktion, Φ , för att relatera varje menings c-struktur med dess respektive f-struktur. Detta – i all korthet – presenterade ramverk har utvidgats och generaliserats till att inkludera även andra sorters lingvistisk information (Kaplan 1987; Halvorsen 1988). Halvorsen & Kaplan (1988) har visat hur den semantiska strukturen kan sättas i korrespondens med såväl c- som f-strukturerna genom mappningar σ respektive σ' .

Denna utvidgade, deskriptiva arkitektur inom LFG tillhandahåller möjligheten att relatera flera strukturnivåer via separata korrespondenser. Genom att introducera en funktion, τ , så skapar man mappningar mellan c- respektive f-strukturerna i ett språk och c-respektive f-strukturerna i ett annat språk, enligt figuren nedan:



Denna metod, menar man, klarar av olika lingvistiska mappningsproblem som ställt till problem för andra metoder. Sålunda klarar man t ex av skillnader i grammatisk funktion, som i följande exempel:

Der Student beantwortet die Frage → L'étudiant répond à la question

Andra exempel som ges inkluderar icke-lokala dependenser mellan det subkategoriserande verbet och en dislokerad fras, som i

*... der Brief, den der Student zu beantworten scheint
 ... la lettre, à laquelle l'étudiant semble répondre
 ... the letter, that the student seemed to answer*

Man klarar med denna metod även av fenomen som skillnader i inbäddning, kommendering med mera.

Metoden har mött kritik från bland andra Sadler & Thompson (1991).

7.6 Quasi Logical Form

Alshawi et al (1991) och Alshawi (1992) redogör för ett översättningssystem som använder sig av så kallade kvasi-logiska former – *Quasi-Logical Forms* – för transfer. En mening i SL analyseras av en modul som benämnes CLE, *Core Language Engine*, som ger meningen en syntaktisk såväl som en semantisk representation i samma (prolog-baserade) formalism. På den kvasi-logiska nivån sker sedan transfer till en motsvarande kvasi-logisk form i TL. Att man valt QLF-nivån för transfer beror på att meningar på den nivån är tillräckligt långt ifrån ytformerna för att lätt låta sig omstruktureras, men samtidigt tillräckligt nära för den lingvistiska, unifikationsbaserade process som används för att skapa QLF:erna. Ambiguiteter löses genom en kontextuell tolkning som ger som resultat en så kallad RQLF, för *Resolved QLF*.

Schematiskt kan arkitekturen åskådliggöras sålunda:

Mening → lingvistisk analys → QLF
QLF → kontextuell tolkning → RQLF

Såväl den lingvistiska som den kontextuella analysen inkluderar ett flertal olika faser, som ortografisk, morfologisk, syntaktisk och semantisk analys. Samtliga dessa analyser bygger på deklarativa bidirektionella regler inom ett unifieringsgrammatiskt paradigm.

Att man valt QLF-nivån för transfer beror på att LF-nivån är för abstrakt för att transfer skall kunna äga rum. Bland annat så har här information som topikalisering gått förlorad. Man vill heller inte använda sig av de 'resolverade' formerna, RQLF:erna, eftersom inte all information om hur de löstes behövs för själva översättningensfasen.

Schematiskt kan själva transferregeln beskrivas sålunda (i prologformalism):

```
trans(<QLF1 subexpression pattern>  
      <Operator>  
      <QLF2 subexpression pattern>).
```

Om operatoren i detta fall utgörs av == så är regeln bidirektionell, medan <= respektive >= anger unidirektionella regler. På grund av att reglerna används icke-deterministiskt och med ren unifiering erhålls monotonicitet på köpet.

Ett system som använder sig av CLE för översättning är SLT-projektet, *Spoken Language Translation*, som sker i samarbete mellan SRI i Menlo Park, SRI i Cambridge (England), Telia Research AB, Haninge. SLT syftar till att skapa system för tal-till-talöversättning mellan engelska och svenska inom flygbokningsdomänen och inkluderar således även moduler för taligenkänning och talsyntes.

8 Utvärdering av system

Ett alldeles särskilt problem utgör själva utvärderingen av hur ett system fungerar. Hur kan man svara på frågan huruvida ett system är "bra" eller inte? För att svara på denna fråga måste man först inse att ett system kan vara 'bra' på flera olika sätt och nivåer. Nedan skall redogöra för några av dessa:

8.1 Organisatoriskt

Hur väl lämpar sig ett system för inkorporering i en redan existerande struktur rent företagsmässigt. Är systemet lätt att använda? För att översättningssystem skall fungera i praktiken krävs det ofta att den textproducerande personalen anpassar sitt språk en smula, genom att t ex undvika frasstrukturer som typiskt leder till strukturell ambiguitet eller dylikt. Det är också viktigt att systemet rent tekniskt går att inlemma i den miljö där det skall användas.

8.2 Hastighet

Denna fråga är något mindre väsentlig än den till en början kan verka. Om ett system kräver *on-line*-interaktion med en mänsklig översättare så är hastighet viktig. I andra fall gör det ingenting om systemet står och tuggar natten igenom och presenterar resultatet nästa morgon.

8.3 Kvalitet

Denna punkt är däremot ytterligt väsentlig för huruvida ett MT-system är intressant eller inte. Problemet att utvärdera översättningar är inte unikt för MT, utan gäller i lika hög grad mänskliga översättare. Man brukar främst iakttaga två parametrar:

Den första rör ett systems **begriplighet**. Ett dokument kan bli obegripligt på flera sätt: grammatiska fel, felöversättningar, oöversatta ord med mera. Även stil kan ibland minska begripligheten hos ett dokument.

Det andra rör systemets **trohet** till originalet. Även om en mening i TL är grammatiskt korrekt och fullt begriplig så är detta tämligen ointressant om den inte återger betydelse i ursprungsmeningen i SL. I normalfallet har det emellertid visat sig att trohet oftast är ett resultat av begriplighet, varför det är intressantare att utvärdera det senare. Om ett dokument är begripligt så är det i regel troget SL.

9 Framtida MT

Huruvida framtida MT någonsin kommer lyckas översätta Shakespeare på ett tillfredställande (eller bra!) sätt är en fråga som få ställer sig. Till yttermera visso är det ytterst få som ens ser någon poäng i att detta skulle ske! Snarare är det så att man anser att översättare kommer att få mera tid över till att översätta litteratur om maskiner tar hand om byråkratisk text, manualer och dylikt. Således är översättning av skönlitteratur inte ens önskvärd.

Vad som däremot är av stort intresse inom framtida MT är inkorporeringen av MT i andra språkteknologiska system. Således kräver ju tal-till-tal-översättning en fungerande MT-modul för att fungera. Vidare så kan talstyrd databassökning ske till databaser på främmande språk om gränssnittet inkluderar en översättande modul.

Så länge tal och språk är och förblir människans viktigaste kommunikationsmedel, och så länge människor talar olika språk kommer det antagligen att finnas intresse för forskning inom MT, av såväl ekonomiska som vetenskapliga skäl.

10 Olika metoder: En sammanställning med bibliografi

Vi har ovan sett exempel på olika strategier för att lösa ett av datorlingvistikens stora problem, tillika ett av de initiala. Här följer en kort sammanfattande lista över de genomgångna metoderna samt en kompletterande litteraturlista.

- **Interlinguala**

Förutom beskrivningen ovan hänvisas till Nirenberg & Goodman (1989); Dorr (1990, 1992).

- **Transfer**

Diverse system beskrivs av Estival et al (1990); Russell et al (1991); Zajac (1989, 1990); Arnold & Sadler (1990); van Noord et al (1990).

- **Shake and Bake**

Förutom Whitelock (1992) och Brew (1992) kan nämnas Beaven (1992).

- **Co-description**

Ytterligare information finns att hämta i Sadler & Arnold (1992); Kaplan & Wedekind (1993); Wedekind & Kaplan (1993) samt Eberle et al (1992).

- **Flexibla metoder**

Termen ”flexibla metoder” har lånats av Louisa Sadler och denoterar metoder som kombinerar en eller flera av övriga här nämnda filosofier, med eller utan tillägg av andra drag. Ett exempel på en sådan metod är bl a abduktions-översättning. Några andra sådana system beskrivs i följande: Tsujii & Fujita (1991); Sanfilippo et al (1992); Copestake (1993).

- **Exempel-baserade**

Förutom redan nämnda verk hänvisas till Sumita et al (1990); Sato (1992);

Observera att de direkt-översättande systemen lyser med sin frånvaro i ovanstående förteckning.

11 Exempel på olika system

Jag skall slutligen presentera några av de översättningssystem som förekommer runt om i världen. Som tidigare nämnts är få system ”renrasiga”, utan utgör oftare en blandning av olika metoder.

11.1 Transformer

- SYSTRAN (Hutchins & Somers 1992; Wilks 1992)

11.2 Transfer

- ARIANE (Vauquois & Boitet 1985)
- MU (Nagao et al 1986)
- SUSY (Maas 1987)
- METAL (Slocum et al 1987)
- TAUM-AVIATHON (Isabella 1987)
- ETAP-2 (Apresian et al 1992)
- LMT (McCord 1989)
- EUROTRA (Arnold 1986; Arnold & des Tombe 1987; Copeland et al 1991)

- CAT-2 (Sharp 1988)
- MIMO (Arnold & Sadler 1990)
- MIMO-2 (van Noord et al 1990)
- ELU (Estival et al 1990)

11.3 Interlingua

- Rosetta (Landsbergen 1987a; 1987b)
 - KBMT (Goodman 1989)
 - DLT (Schubert 1992)
-

12 Referenser och bibliografi

- Alshawi, Hiyan. 1992. *The Core Language Engine*. MIT Press, Cambridge, Massachusetts.
- Alshawi, Hiyan, David Carter, Manny Rayner, Björn Gambäck. 1991. Translation by Quasi-Logical Form Transfer. *Proceedings of the 29th ACL*, Berkeley.
- Apresian, Jurij, Igor Boguslavskij, Leonid Iomdin, Alexandre Lazurskij, Vladimir Sannikov, Leonid Tsinman. 1992. ETAP-2: The Linguistics of a Machine Translation System. *Meta* XXXVIII(4):97–112.
- Arnold, Doug & Lousia Sadler. 1990. The Theoretical Basis of MiMo. *Machine Translation*, Vol 5, pp. 195–222.
- Arnold, D, L Balkan, R Lee Humphreys, S Meijer, L Sadler. 1994. *Machine Translation – An Introductory Guide*. Blackwell.
- Arnold, D J och Louis des Tombe. 1987. Basic Theory and methodology in Eurotra. Sergei Nirenburg (red): *Machine Translation: Theoretical and Methodological Issues*. Sid 114–135. Cambridge University Press, Cambridge.
- Arnold, D J. 1986. Eurotra. A European perspective on MT. *Proceedings of the IEEE* 74(7):979–992.
- Beaven, John. 1992. Shake-and-Bake Machine Translation. *Proceedings of the 15th international conference on computational linguistics (Coling) 92*, Nantes.
- Brew, C H. 1992. Letting the Cat out of the Bag: Generation for Shake-and-Bake MT. 1992. *Proceedings of the 15th international conference on computational linguistics (Coling) 92*, Nantes.
- Copeland, C, J Durand, S Krauwer, B Maegaard (reds).1991. *The Eurotra Linguistic Specifications. Vol 3 of Studies in Machine Translation and Natural Language Processing. Office for Official Publications of the Commission of the European Community*. Luxemburg.
- Copestake, A & A Sanfilippo. 1993. Multilingual Lexical Representation. *Proceedings of the AAAI-93 Spring Symposium Series: Building Lexicons for Machine Translation*, Stanford, California.
- Dorr, Bonnie. 1992. Parameterization of the Interlingua in Machine Translation. 1992. *Proceedings of the 15th international conference on computational linguistics (Coling) 92*, Nantes.
- Dorr, Bonnie. 1990. Cross-Linguistic Approach to Translation. *Proceedings of the 3rd International Conference on Theoretical and Methodological Issues in Machine Translation*, Austin, Texas.
- Eberle, Kurt, Walter Kasper, Christian Rohrer. 1992. Contextual Constraints for MT. *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation*, Montreal, Canada.
- Estival, Dominique, Afzal Ballim, Graham Russell, Susan Warwick. 1990. A Syntax and Semantics for Feature-Structure Transfer. *Proceedings of the 3rd International Conference on Theoretical and Methodological Issues in Machine Translation*, Austin, Texas.
- van Eynde, Frank. 1993. Machine Translation and Linguistic Motivation. Frank van Eynde (red) *Linguistic Issues in Machine Translation*. Sid 1–43. Pinter Publishers.
- Fillmore, Charles. 1968. The Case for Case. I Bach, E & R Harms (red): *Universals in Linguistic Theory*. Sid 1–90. Holt, Rinehart and Winston, New York.
- Fodor, Jerry. 1977: *Semantics: theories of meaning in generative grammar*. Cromwell, New York.
- Furuse, Osamu & Hitoshi Iida. 1992. An Example-based Method for Transfer-Driven Machine Translation. *Proceedings of TMI14*, Montreal.

- Goodman, K. (red). 1989. *Machine Translation*. Vol 4, 1989, Numbers 1 and 2: special issues on knowledge-based MT, Parts I and II.
- Halvorsen, Per-Kristian. 1988. Situation Semantics and Semantic Interpretation in Constraint-Based Grammars. *Proceedings of the International Conference on Fifth Generation Computer Systems*, Tokyo.
- Halvorsen, Per-Kristian & Ronald Kaplan. 1988. Projections and Linguistic Description. *Proceedings of the International Conference on Fifth Generation Computer Systems*, Tokyo.
- Hobbs, Jerry & Megumi Kameyana. 1990. Translation by Abduction. *Proceedings of Coling 90*, Helsinki.
- Hutchins, W John & Harold L Somers. 1992. *An Introduction to Machine Translation*. Kapitel 6, sid 99–129. Academic Press.
- Isabelle, P. 1987. Machine Translation at the TAUM Group. Margaret King (red): *Machine Translation: The state of the art. Proceedings of the Third Lugano Tutorial*. Edinburgh University Press, Edinburgh.
- Kaplan, Ronald. 1987. Three Seductions of Computational Psycholinguistics. Peter Whitelock et al (red) *Linguistic Theory and Computer Applications*. Sid 149–188. Academic Press.
- Kaplan, Ronald & Jürgen Wedekind. 1993. Restriction and Correspondence-Based Translation. *Proceedings of the 6th EACL*, Utrecht, Holland.
- Kaplan, Ronald & Joan Bresnan. 1982. Lexical Functional Grammar: A Formal System for Grammatical Representation. Joan Bresnan (red): *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, Massachusetts.
- Kaplan, Ronald, Klaus Netter, Jürgen Wedekind, Annie Zaenen. 1989. Translation by Structural Correspondences. *Proceedings of the Fourth Conference of EACL*, Manchester.
- Landsbergen, Jan. 1987. Isomorphic grammars and their use in the Rosetta translation system. Margaret King (red): *Machine Translation: The state of the art*. Edinburgh University Press, Edinburgh.
- Maas, Hans-Dieter. 1987. The MT System SUSY. Margaret King (red): *Machine Translation: The state of the art. Proceedings of the Third Lugano Tutorial*. Sid. 209–246. Edinburgh University Press, Edinburgh.
- McCord, Michael. 1989. Design of LMT: A Prolog-Based Machine Translation System. *Computational Linguistics* 15(1):33–52.
- Nagao, Makato. 1984. A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle. A Elithorn och R Banerji (red), *Artificial and Human Intelligence*. Sid 173–180. North-Holland.
- Nagao, Makato, Jun-ichi Tsujii och Jun-ichi Nakamura. 1986. Machine Translation from Japanese into English. *Proceedings of the IEEE*, 74(7):993–1012, July 1986.
- Nirenberg, S & Ken Goodman (red). 1989. *Machine Translation. Special Issue on Knowledge-Based MT*, Volume 4, numbers 1 and 2.
- van Noord, Gertjan, Joke Dorrepaal, Pim van der Eijk, Maria Florenza, Louis des Tombes. 1990. The MiMo2 Research System. *Proceedings of the 3rd International Conference on Theoretical and Methodological Issues in Machine Translation*, Austin, Texas.
- Russell, Graham, Afzal Ballim, Dominique Estival, Susan Warwick. 1991. A Language for the Statement of Binary Relations over Feature Structures. *Proceedings of the Fifth Conference of EACL*, Berlin.
- Sadler, Louisa. 1989. *Working with Analogical Semantics*. Foris Publications.

- Sadler, Louisa & Doug Arnold. 1992. A Constraint-Based Approach to Translating Anaphoric Dependencies. *Proceedings of the 15th international conference on computational linguistics (Coling) 92*, Nantes.
- Sadler, Louisa & Henry S Thompson. Structural Non-Correspondence in Translation. 1991. I *Proceedings of the Fifth Conference of EACL*, Berlin.
- Sanfilippo, Antonio, Ted Briscoe, Ann Copestake, M A Marti, M Taule, A Alonge. 1992. Translation Equivalence and Lexicalisation in the ACQUILEX LKB. *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation*, Montreal.
- Sato, S. 1992. CTM: An Example-Based Translation Aid System. *Proceedings of the 15th international conference on computational linguistics (Coling) 92*, Nantes.
- Sato, S & M Nagao. 1990. Toward Memory-Based Translation. *Proceedings of Coling 90*.
- Sharp, Randall. 1988. CAT-2 – implementing a formalism for multi-lingual MT. *Proceedings of the 2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*. Sid 76–87. Carnegie Mellon University, Pittsburgh.
- Slocum, J, W S Bennett, J Bear, M Morgan, R Root. 1987. Metal: The CRC Machine Translation System. Margaret King (red): *Machine Translation Today*. Sid 319–350. Edinburgh University Press, Edinburgh.
- Schubert, Klaus. 1992. Esperanto as an intermediate language for Machine Translation. John Newton (red): *Computers in Translation: A Practical Appraisal*. Sid 79–85. Routledge, London.
- Sumita, E, H Iida. 1991. Experiments and Prospects of Example-Based Machine Translation. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*.
- Sumita, E, H Iid, H Kohyama. 1990. Translating with Examples: A New Approach to Machine Translation. *Proceedings of the 3rd International Conference on Theoretical and Methodological Issues in Machine Translation*, Austin, Texas.
- Tsujii, Jun-ich & Kimikazu Fujita. 1990. Lexical Transfer Based on Bilingual Sign: Towards Interaction During Transfer. 1991. *Proceedings of the Fifth Conference of EACL*, Berlin.
- Vauquois, Bernard och Christian Boitet. 1985. Automated Translation at Grenoble University. *Computational Linguistics* 11(1):20–36.
- Wedekind, Jürgen & Ronald Kaplan. 1993. Type-Driven Semantic Interpretation of f-structures. *Proceedings of the 6th EACL*, Utrecht, Holland.
- Wilks, Yorick. 1992. Systran: It obviously works, but how can it be improved? John Newton (red): *Computers in Translation: A Practical Appraisal*. Sid 116–188. Routledge, London.
- Whitelock, P. Shake-and-Bake Translation. 1992. *Proceedings of Coling 92*, Nantes.
- Zajac, Remi. 1990. A Relational Approach to Translation. *Proceedings of the 3rd International Conference on Theoretical and Methodological Issues in Machine Translation*, Austin, Texas.
- Zajac, Remi. 1989. A Transfer Model Using a Typed Feature Structure Rewriting System with Inheritance. *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, University of British Columbia, Vancouver, Canada.
-