

STEFÁN BRIEM

Maskinoversættelse fra esperanto til islandsk

Abstract

Machine Translation from Esperanto to Icelandic

It has become evident that Esperanto, due to its logicity and regularity, is well suited as an intermediate language in machine translations between ethnic languages. The project presented here is intended to be a contribution to a later inclusion of Icelandic in a larger multilingual system for machine translation, where Esperanto is used as an intermediate language. In fact such a multilingual machine translation system, DLT (Distributed Language Translation), is being developed at BSO/Research in Utrecht (Netherlands).

My project started in the year 1981. It was a spare time activity exclusively on a private basis until two years ago when it received public support in Iceland. This is a machine translation system applicable as an instrument for studying various ideas relating to machine translation from Esperanto to Icelandic.

An effective machine translation system must be capable of coping with unexpected words, i.a. words created by spontaneous need in the daily use of the language. Those new creations follow rules and conventions that can be different from one language to another. This theme, that involves automatic word creation, is discussed with respect to Esperanto and Icelandic. The derivation of words from words of different word classes by changing the endings of the words, is a common feature of Esperanto and Icelandic. In cases where this happens in a similar and systematic way in both languages, automatic word creation is feasible. A noticeable characteristic of Esperanto is the extensive use of affixes, both prefixes and suffixes, in order to modify the meaning of words and to create new words. In Icelandic affixes are used less frequently. This means that when translating between Esperanto and Icelandic sometimes suffixed words in Esperanto correspond to different word constructions in Icelandic.

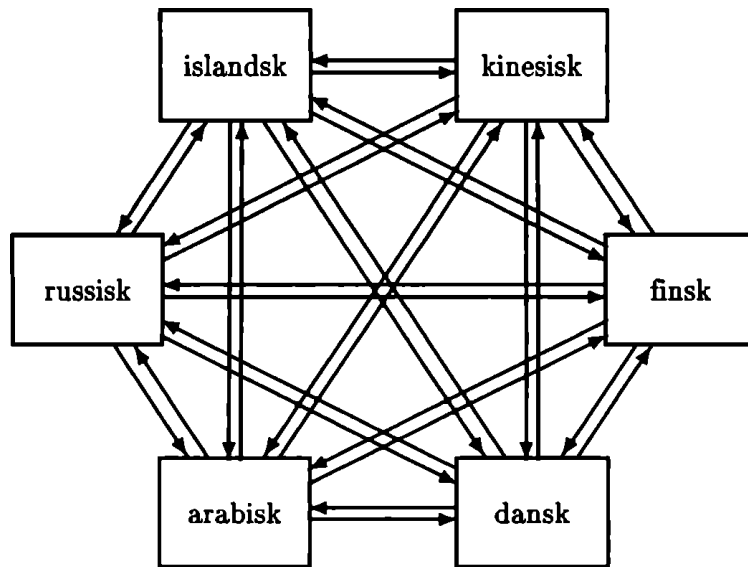
1 Indledning

I de sidste otte år har jeg beskæftiget mig med maskinoversættelse fra esperanto til islandsk, ganske vist ikke uafbrudt. De første 6 år var det udelukkende

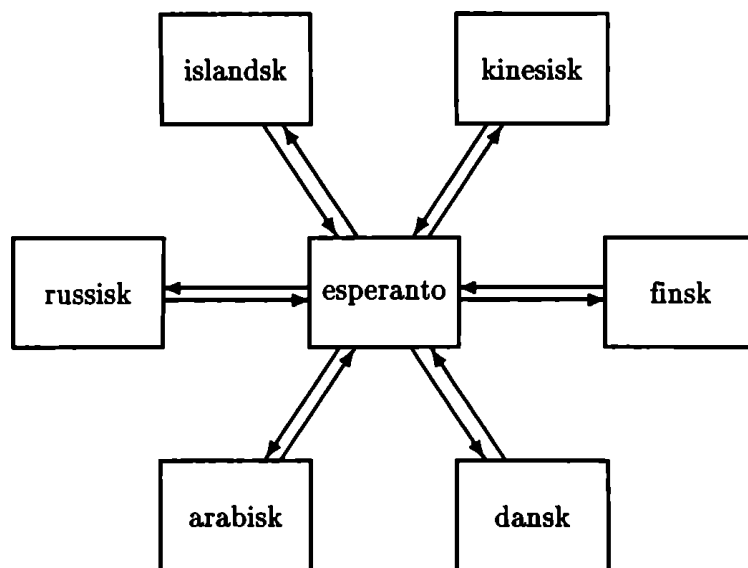
fritidsbeskæftigelse helt på egne vegne. Men i den sidste tid har mit projekt mødt stadig voksende interesse og har fået offentlig støtte.

Jeg vil begynde med at placere mit projekt i større sammenhæng.

Hvis man betragter maskinoversættelse parvis mellem mange sprog så kunne den tænkes at foregå direkte mellem hvert sprogpar.



En anden mulighed er at bruge et fælles mellemsprog, hvor f.eks. esperanto tjener som mellemsprog.



Valget af esperanto som mellemsprog er ikke tilfældigt, men baseres på, at esperanto er et særdeles logisk sprog med regelmæssig grammatik, hvorfor det

er velegnet til datalingvistisk brug. Det er netop denne idé, som jeg har sluttet mig til, da jeg har valgt at beskæftige mig med maskinoversættelse fra esperanto til islandsk. Grunden til at jeg hidtil har begrænset mig til oversættelse i denne retning, men ikke fra islandsk til esperanto, er simpelthen den at parsning af esperanto er meget nemmere end parsning af andre sprog, og der er jo alligevel problemer nok at glæde sig over.

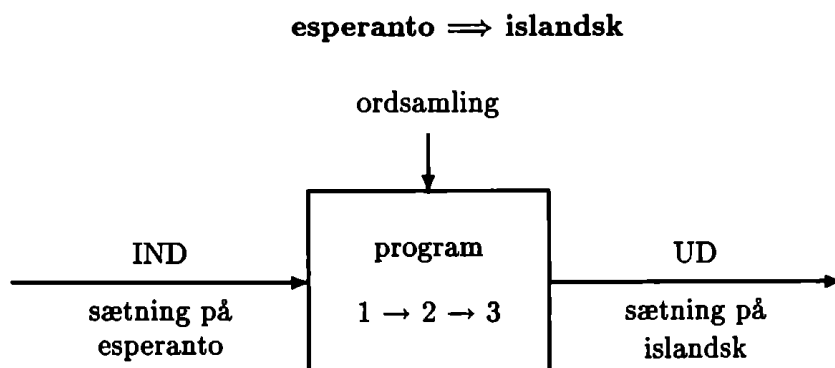
Med henblik på denne grundlæggende idé går mit projekt i takt med det kendte DLT-projekt for multispøglig maskinoversættelse, hvor man også har valgt esperanto som mellemsprog.

I tidens løb har jeg udviklet et oversættelsesprogram sammen med en ordsamling på godt 6.000 ord. Programmet er i stand til at gengive esperanto-tekst på islandsk, således at de islandske ord har den ønskede bøjningsform og sætningernes struktur er ændret i retning af korrekt islandsk syntaks.

Jeg har fundet det nyttigt at have et oversættelsesprogram, som fungerer. Selv om programmet meget lidt berører semantik så kan det bruges til at studere forskellige problemer ved maskinoversættelse, især syntaktiske og morfologiske problemer. Man kan eksperimentere med idéer og mulige løsninger til disse problemer ved at ændre lidt på program og ordsamling og, hvad der kan være meget opmuntrende, hurtigt se et resultat på dataskærm eller på papir.

2 Programmets funktion

Hvordan fungerer oversættelsesprogrammet?



2.1 Etaper:

1. Analyse
2. Oversættelse ord for ord
3. Strukturændringer

Programmet arbejder med én sætning ad gangen og går fra det enkle til det mere komplicerede. Inddata er altså en sætning på esperanto. Programmets første operation er at dele sætningen op i enkelte ord. Den anden operation er at slå op i ordregistre som giver for hvert enkelt esperanto-tekstord det tilsvarende

islandske ord sammen med morfologiske oplysninger. Den tredje operation består af strukturændringer i henhold til forskellen på esperanto og islandsk. Til slut afleverer programmet som uddata en islandsk sætning med næsten islandsk syntaks og forhåbentlig nogenlunde samme mening som den oprindelige sætning på esperanto har.

Det, som jeg især vil diskutere ved denne lejlighed, er hvilke udveje man har når et esperanto-ord ikke findes direkte i ordsamlingen. Løsningen af dette problem indebærer automatisk orddannelse.

3 Ordsamlingens inddeling

På grund af esperantos simple grammatik afdækkes et tekstords ordklasse og bøjningsform på grundlag af ordets endelse alene, uafhængigt af de andre ord i samme sætning. Jeg har derfor fundet det nyttigt at dele ordsamlingen op i fem ordregistre som vist her.

esperanto \Rightarrow islandsk

	endelser
1. Substantiver	o <i>oj on ojn</i>
2. Adjektiver	a <i>aj an ajn</i>
3. Adverbier	e <i>en</i>
4. Verber	i <i>as is os us u</i>
5. Småord	(pronominer, præpositioner, konjunktioner, adverbier, numeralier o.fl.)

Register nummer 5 indeholder en lukket klasse af ca. 200 ord. De andre registre indeholder åbne ordklasser. Hvis et esperanto-ord ikke findes i register 5 så tilhører det med sikkerhed én bestemt af de fire andre registres ordklasser. Såfremt ordet er til stede i ordsamlingen, finder programmet hurtigt frem til det. Men hvad hvis ordet alligevel ikke findes? Og det vil man jo ikke kunne undgå i tilfælde af nydannelser.

4 Automatisk orddannelse

Man kunne selvfølgelig simpelthen give op og lade være med at oversætte ordet. I tilfælde af esperanto kunne man alligevel holde rede på syntaksen, idet ordets ordklasse og bøjningsform er kendt fra endelsen, selv om stammen måske er ukendt. Men der er også den mulighed at ordets stamme eller bestanddele faktisk er til stede selv om ordet ikke findes direkte i registret for den rigtige ordklasse.

Man kan dele de ukendte esperanto-ord i fem tilfælde:

1. Sammensatte ord
2. Ord afledte fra andre ordklasser

3. Ord med affikser
4. Kombinationer af 1-3
5. Andre ord

Af disse tilfælde har jeg især eksperimenteret med 2 til 4 og i den anledning inkluderet nogen orddannelsesregler i mit oversættelsesprogram.

Lad os først betragte de andre tilfælde.

Sammensatte ord er mindre almindelige i esperanto end i islandsk. I de fleste tilfælde kan man nok automatisk oversætte et sammensat ord direkte fra esperanto til islandsk, d.v.s. man finder frem til de enkelte stammers oversættelser og danner af dem et sammensat ord på islandsk.

Hvad 5. tilfælde angår, *andre ord*, så drejer det sig om ord som ikke kan oversættes i de omgivelser. Man kan simpelthen gengive ordet uden oversættelse i den islandske tekst. Man kan jo alligevel benytte sig af kendskabet til ordets ordklasse og bøjningsform til at bestemme den islandske sætnings struktur.

4.1 Afledning fra en anden ordklasse

Jeg vil nu vise et par eksempler på automatisk orddannelse ved afledning fra andre ordklasser.

I esperanto er denne type af orddannelse meget almindelig, meget mere end i islandsk og dansk f.eks.

Eksempel 1. adverbium dannet af adjektiv

adjektiv:	<i>feliĉa</i>	<i>hamingjusamur</i>	lykkelig
adverbium:	<i>feliĉe</i>	<i>hamingjusamlega</i>	lykkeligt

Når programmet skal oversætte adverbiet *feliĉe* til islandsk, så vil dette ord ikke findes i samlingen af adverbier. Programmet vil da søge i samlingen af adjektiver for adjektivet *feliĉa*, som jo findes der. Resultatet bliver det islandske adverbium *hamingjusamlega* afledt af det islandske adjektiv *hamingjusamur*.

Det andet eksempel drejer sig om et adjektiv, eller snarere om et participium, som behandles som et adjektiv.

Eksempel 2. adjektiv (participium) dannet af verbum

verbum:	<i>kuri</i>	<i>hlaupa</i>	løbe
adjektiv:	<i>kuranta</i>	<i>hlaupandi</i>	løbende

Ordet *kuranta* findes ikke i samlingen af adjektiver. Da ordet ender på *anta* er det muligvis præsens participium af et verbum. Programmet vil derfor søge i samlingen af verber for verbet *kuri*. Resultatet er positivt og programmet danner automatisk ordet *hlaupandi* ved afledning af det islandske verbum *hlaupa*.

De to eksempler kan også kombineres og blive til en dobbeltafledning.

Eksempel 2 + 1. adverbium dannet af verbum

adverbium: kurante hlaupandi løbende

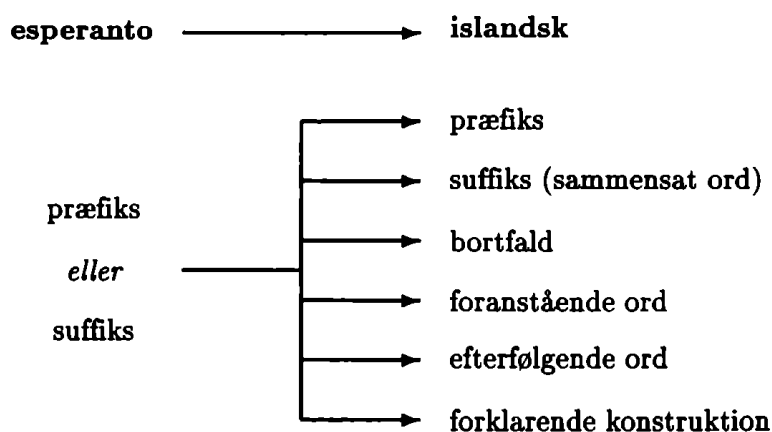
Adverbiet *kurante* oversættes ved at programmet først søger forgæves efter adjektivet *kuranta* og derefter finder frem til verbet *kuri*.

4.2 Esperanto-ord med affikser

En anden type af orddannelser på esperanto går ud på udstrakt brug af affikser, både præfikser og suffikser.

Når esperanto-ord med affikser skal kunne oversættes til islandsk, kan de mest almindelige med fordel inkluderes i ordsamlingerne.

For de andre affiksede ords vedkommende kan man prøve at vælge en af de oversættelsesmetoder som antydes nedenfor.



Både i tilfælde af et præfiks og et suffiks kommer alle disse muligheder i betragtning. Man kan vise eksempler på alle mulighederne, måske med én undtagelse. Om lidt vil vi se nogle eksempler.

Men først skal vi se på en særpræget brug af præfikser på esperanto. Det drejer sig om konstruktionstvillinger, hvor en præposition kan placeres som præfiks på et verbum med ingen eller minimal ændring af mening.

4.2.1 Strukturændring på esperanto

esperanto: Mi parolas pri la libro.
 islandsk: Ég tala um bókina.
 dansk: Jeg taler om bogen.

Den samme mening kan ytres ved sætningen:

esperanto: Mi priparolas la libron.

Når den sidste sætning skal oversættes til islandsk bliver dens struktur først ændret ved at præfikset **pri** laves om til præpositionen **pri**, der placeres efter verbet, samtidig med at det akkusative **n** fjernes fra substantivet.

Mi **pri**parolas la libro**n**.

4.2.2 Oversættelse af præfikser

Her får vi eksempler på 5 muligheder for oversættelse af ord med præfikser.

1.	miskompreni	misskilja	misforstå
2.	senlabora	atvinnulaus	arbejdsløs
3.	kamaradoj	félagar	kammerater
	gekamaradoj	félagar	kammerater (af begge køn)
4.	eksministro	fyrirverandi ráðherra	forhenværende minister
5.	foriri	fara burt	gå bort

4.2.3 Oversættelse af suffikser

Og nu over til 5 eksempler på oversættelse af ord med suffikser.

1.	leono	ljón	løve
	leonino	kvenljón	hunløve
2.	grupo	hópur	gruppe
	grupestro	hópstjóri	gruppebestyrer
3.	paroli	tala	tale
	paroladi	tala	tale (vedvarende)
4.	homo	maður	menneske
	homaĉo	lítilfjörlegur maður	usselt menneske
5.	skribi	skrifa	skrive
	skribema	hneigður fyrir að skrifa	tilbøjelig til at skrive

4.3 Kombinationer

Nu har vi betragtet nogle eksempler på tilfælde nr. 2 og 3 af ukendte esperanto-ord. Ifølge tilfælde nr. 4 kan man også have kombinationer af 1, 2 og 3, og den slags kombinationer er endog ret almindelige i esperanto. Når ord af denne art skal oversættes, kan det være afgørende for resultatets kvalitet i hvilken rækkefølge ordets struktur analyseres. Den mest gunstige rækkefølge vil selvfølgelig ikke blive den samme for de forskellige ordklasser. I denne henseende er det min filosofi, at man ved hjælp af et simpelt oversættelsesprogram som mit kan afprøve de forskellige muligheder på en naturlig tekst og derved finde frem til den optimale rækkefølge for hver ordklasse.

5 Forøgelse af ordsamlingen

Den begrænsede størrelse af ordsamling virker ofte utilstrækkelig, når man vil eksperimentere med en naturlig tekst. Derfor arbejder jeg nu på en forøgelse af antal ord fra godt 6.000 til knap 24.000, d.v.s. næsten firedobling. Dette sideprojekt, som nu er i den sidste fase, består i ud fra en temmelig stor islandsk-esperanto ordbog (Skafthell 1965) i trykt form at fremstille en esperanto-islandsk ordsamling i datamatlæsbar form. En esperanto-islandsk ordbog af tilsvarende omfang ville selvfølgelig have været nemmere at arbejde med, men den eksisterer desværre ikke. Konverteringen er tilvejebragt til dels ad mekanisk vej, idet den består af følgende etaper:

- Optisk karakterrekognition (ved hjælp af Kurzweil-maskine)
- Maskinstøttet konvertering fra islandsk-esperanto til esperanto-islandsk
- Manuel bearbejdelse

Litteratur

Briem, Stefán. 1988. *Vélrænar tungumálabyðingar*. Projektrapport. Reykjavík.

Skafthell, Baldvin B. 1965. *Íslensk-esperanto orðabók*. Samband íslenzkra esperantista. Reykjavík.