

A general programme for mechanical translation between any two languages via an algebraic interlingua

by R.H. Richens

It has become clear that the amount of lexical and syntactical analysis required to produce a smooth and idiomatic mechanical translation from any base language into any target language is very great. In this paper the possibilities of mechanical translation via an algebraic interlingua are explored. With this approach, only one machine programme is envisaged for translation between any two languages. A mechanical dictionary is required for each base and each target language.

The interlingua

The algebraic interlingua under investigation is ideographic and constructed so as to represent the ideas of any passage in a base or target language divested of all lexical and syntactical peculiarities; for which reason it is called Nude.

The words in Nude are constructed of some fifty elements (Roman letters and numerals, capital and lower case letters being regarded as different symbols), each of which denotes some basic idea such as plurality, plant or negation. A word in Nude may consist of one letter only; the more complex a notion, the more elements are required. Superior punctuation, apostrophes and quotes, are used as brackets within a word.

Each word in Nude is regarded as a relation, either 0-ad, 1-ad or 2-ad.

Common nouns, usually correspond to 0-ads, e.g. Xp'CL = seed.

Adjectives and intransitive verbs, e.g. red or blush, are usually 1-ad; these are preceded by a point, e.g. .Pz = one.

Transitive verbs exemplify 2-ads; these are preceded by a colon, e.g. :B''f'CL = ripen.

The word qualified by a 1-ad precedes it, e.g. Xp'CL.Pz = one seed.

The words linked by a 2-ad precede it and are separated by a comma, e.g. p,Xp'CL.Pz:B''f'CL = the plant is ripening a single seed.

Coordinate conjunction, in which two words or word groups are qualified in the same way by the same 1-ad or 2-ad, are joined by a hyphen.

Translation from Japanese to Nude

The base passage is:-

KETSU SAKU HO GO
HEI ni ICHI SAKU TO² ri
SHU SHI RYU SU² ha
KO HAI JI KI ni yo tsu te
I ru

(1.1)

Words in capitals represent the on pronunciation of the Chinese ideographs in the text; lower case letters are used for syllables in hiragana script; the superscripts are used to designate derivative (abbreviated) forms of the ideographs. The above reading does not, of course, represent the pronunciation of the passage.

The first operation, assuming that the difficult process of character recognition has been accomplished, is to replace all derivative ideographs by the standard forms. Thus:-

KETSU SAKU HO GO
HEI ni ICHI SAKU TO ri
SHU SHI RYU SU ha
KO HAI JI KI ni yo tsu te
I ru

(1.2)

Next, the text has to be decomposed into semantically significant chunks by matching against the Japanese-Nude dictionary. In Japanese, there is no indication as to where words or chunks begin or end. It is therefore necessary to start at the beginning of the passage, match the longest sequence, headed by the first character or syllable, that the dictionary contains, and then proceed to the next sequence. In the passage quoted, KETSU is the first chunk, SAKU the second, and HO GO the third.

When a match is made, the following entries in the mechanical dictionary, corresponding to each chunk of the base passage, are extracted:-

- {1} The corresponding word or words in Nude.
- {2} Category information.
- (3) Word Class.

By category is meant conjugation, declension and any other indication of morphological status. Word class is an extension of part of speech. There is no necessity for any cross correspondence between the word classes used for any two languages.

The output at this stage for the passage being considered is:-

:KETSU x:^u'CN N * SAKU xp'"CL'HI'Q N * HO GO nz'P'"P100 N *
HEI ni C * .ICHI Pz Num * SAKU xp'"CL'HI'Q N * TO ri Post I *
SHU SH:^u Ap'"CL N * RYU Nz'^o N * SU P * ha Post III
KO HAI CJ'Sz'"pN / :CL'Sz N * JI KI T N * ni yo tsu te * :C3 Post II *
.I Sz N ru * ru Vt ru
(1.3)

The asterisks mark the articulations of the passage. For ease in exposition, the original chunk together with its translation in Nude will be replaced, as far as feasible, by a catch word in italics. Thus 1.3 can be represented as:-

:tie N * capsule N * percentage N *
C * one Num * capsule N * Post I *
seed N * grain N * number N * Post III *
hybrid N / :hybridize N * time N * :according Post II *
.different N ru * Vt ru
(1.4)

The next stage is the cyclical reduction of the syntax to Nude syntax. This is accomplished by matching word class sequences against the word class sequence entries in the Japanese-Nude mechanical dictionary, following the procedure described by Richens (in the press) at the 1955 Symposium of the Cambridge Language Research Unit. Each matched sequence is replaced by a single word class, and at the same time the order of the words may be changed, and punctuation and additional words may be inserted, instruction for these operations being extracted from the word-class-sequence dictionary entries. The order in which the word class sequences are dealt with is important and it is not suggested that the order adopted here is necessarily the best.

In the present instance, qualifier-noun sequences are taken first. The first reduction is:-

:tie N * capsule N * percentage N *
C * (capsule .one) N * Post I *
seed N * grain N * number N * Post III *
hybrid / :hybridize N * time N * :according Post II *
.different N ru * Vt ru
(1.5)

Next noun - noun and noun - verb transformer sequences are treated, giving, firstly:-

(capsule :tie) N * percentage N *
C * (capsule .one) N * Post I *
(grain N .seed) N * number N * Post III *
(time .hybrid / ,:hybridize) N * :according Post II *
(.different ru ru)V
(1.6)

and secondly:-

(percentage .capsule :tie) N *
C * (capsule .one) N * Post I *
(number .grain .seed) N * Post III *
(time .hybrid /,:hybridize) N * :according Post II *
(.different ru ru) V (1.7)

Postpositional phrases are reduced next:-

(percentage .capsule :tie) N *
C * (number .grain .seed
.capsule .one) N * Post III *
(time .hybrid /,:hybridize :according) Av *
(.different ru ru) V (1.8)

and then coordinate nouns:-

(percentage . capsule . tie
- number .grain .seed
.capsule .one) N * Post III *
(time .hybrid /,:hybridize : according) Av *
(.different ru ru) V (1.9)

Adverbial phrases are now absorbed into their verbs:-

(percentage .capsule .tie
- number .grain .seed
.capsule .one) N * Post III *
(.different ru ru
,time .hybrid /,:hybridize :according) V (1.10)

and, having obtained the word class sequence N Post III .V, the final stage in reduction can be effected:-

(percentage .capsule .tie
- number .grain .seed
.capsule .one
.different ru ru
,time .hybrid /,:hybridize :according) (1.11)

It should be noted at this point that had the word class sequence been of the form N Post III :V, the reduction would have been more difficult since though Post III, the postposition ha (pronounced wa), usually follows the subject of a verb, it can also follow the object.

Before completing this stage of the translation, various interactions have to be dealt with. These are ascertained by consulting the relevant part of the Japanese-Nude dictionary in the way described by Richens and Booth (1955) and amplified by Richens (loc. cit.).

The only chunk-category interaction is I ru ... ru , ru being the conclusive ending of the ru conjugation. There is no additional information provided by this inflection, so the translation now stands as:-

percentage .capsule .tie
- number .grain .seed
.capsule .one
.different
,time .hybrid /,:hybridize :according (1.12)

There is one interesting chunk-semantic interaction in capsule .tie, which stands for SAKU Xp'CL'HI'Q .KETSU xu^oCHI. The interaction is Xp'CL .KETSU, which generates B''f'CL, meaning mature. In other words, when the ideograph KETSU is combined with characters denoting fruits or seeds, it means mature instead of tie. When this

substitution is made, the translation into Nudc is completed and the Japanese chunks can be dropped, giving:-

nz'P''P100 .,Xp'CL'HI'Q :B''f'CL
- P .Nz'c .Xp'CL
.Xp'CL'HI'Q .Pz
.Sz
,T .CL'Sz' p/,:CL'Sz :C3

(1.13)

Whether KO HAI means hybrid (CL'Sz' p) or hybridization (CL'Sz) is uncertain from the passage quoted and cannot even be decided from the immediate context. Both readings should therefore be given. For convenience, however, hybridization (CL'Sz) will be taken as the reading for the following translations.

Translation from Nudc to English

The procedure described above is now repeated. Consultation with the Nudc-English mechanical dictionary provides:-

percentage nz'P''P100 N ab * .,capsule Xp'CL'HI'Q N * :matur B''f'CL V e *
- number P N ab * grain N c N * .seed Xp'CL N *
.capsule Xp'CL'HI'Q N * .one Pz Aj sing *
.different Sz Aj *
,time T N ab ,:hybridiz CL'Sz V e * :according to C3 Pre

(2.1)

Using the same convention of catch words as before, the above can be represented more shortly as:-

percentage N ab * .,capsule N * :mature V *
- number N ab * .grain N * .seed N *
.capsule N * .one Aj sing *
.different Aj *
,time N ab * ,:hybridize * :according Pre

(2.2)

Adjective-noun sequences are dealt with first. This involves reversal of word order and, since the adjective concerned is a singularizing form, transformation of the number-neutral noun into the singular. Thus:-

percentage N ab * .,capsule N * :mature V *
- number N ab * .grain N * .seed N *
.one capsule N *
.different Aj *
,time N ab * ,:hybridize * :according Pre

(2.3)

The next three stages deal with noun-noun and noun-verb sequences. The approximate method of treating neutral common nouns as plural and neutral abstract nouns as singular is adopted. The definite article is inserted before singular nouns, though much more complex rules are required for general use.

percentage N ab * (nature pastpt capsule plur) N plur *
- (the number of grain plur) N * .seed N *
.one capsule N *
.different Aj *
,(the time of hybridize prept) N :according Pre

(2.4)

(the percentage of mature pastpt capsule plur) N *
- (the number of grain plur of seed plur) N *
.one capsule N *
.different Aj *
,(the time of hybridize prept) N * :according Pre

(2.5)

(the percentage of mature pastpt capsule plur) N *
- (the number of grain plur of seed plur
of one capsule) N *
.different Aj *
,(the time of hybridize prept) N * :according Pre

(2.6)

Prepositional phrases are dealt with now:-

- (the percentage of mature pastpt capsule plur) N *
-(the number of grain plur of seed plur
of one capsule) N *
.different Aj *
.according the time of hybridize prept) Av

(2.7)

and then coordinate nouns:-

- (the percentage of mature pastpt capsule plur
and the number of grain plur of seed plur
of one capsule) N plur *
.different Aj *
.according the time of hybridize prept) Av

(2.8)

The word class sequence is now N plur .Aj .Av . From this the final syntactic arrangement is obtained:-

- (the percentage of mature pastpt capsule plur
and the number of grain plur of seed plur
of one capsule
be irr plur different
according the time of hybridize prept)

(2.9)

Chunk-category interactions are now ascertained in the Nude-English dictionary to obtain the inflections for the nouns and verbs, and the final result is:-

- the percentage of matur -ed capsule -s
and the number of grain -s of seed -s
of one capsule
are different
according to the time of hybridiz -ing

(2.10)

Translation from Nude to German

The initial output from the Nude-German dictionary is:-

- Prozentsatz nz'P' P100 N ab mas *., Kapsel Xp"CL'HI'Q N fem * :reif B"f' CL V *
-Zahl P N ab fem * .Gran Mz'c N neut * .Same Xp'CL N fem *
.Kapsel Xp"CL'HI'Q N fem * ein Pz Aj sing *
.verschieden Sz Aj *
,Zeit T N ab fem * ,: bastardier CL'Sz V * :gemäss C3 Pre gen

(3.1)

or, using the catch-word convention:-

- percentage N ab mas * .,capsule N fem * :mature V *
number N ab fem * .grain N neut * .seed N fem *
.capsule N fem * .one Aj sing *
.different Aj *
,time N ab fem * ,:hybridize V * :according Pre gen

(3.2)

The succeeding reductions are parallel to these for the Nude-English translation but the expressions derived are more complex in view of the much more highly inflected nouns, adjectives and verbs and the inflected article:-

- percentage N ab mas * .,capsule N fem * :mature V *
number N ab fem * .grain N neut * .seed N fem *
.(one fem capsule fem) N fem 12 *
.different Aj *
,time N ab fem * ,:hybridize V * :according Pre gen

(3.3)

percentage N ab mas * .(der fem plur mature pastpt fem plur capsule fem
plur) N fem 123 *
-(der fem number fem der gen plur mas grain gen neut plur) N fem 12 *
.seed N fem *
.one fem capsule fem) N fem 12 *
.different Aj *
,(der fem time fem der gen neut hybridize inf gen) N fem 12 * :according
Pre gen

(3.4)

(der mas percentage mas der gen fem plur mature postpt gen fem plur capsule
gen fem plur) N mas 12 *
-(der fem number fem der gen plur mas grain gen neut plur der gen fem plur
seed gen fem plur) *
.one fem capsule fem) N fem 12 *
.different Aj *
,(der fem time fem der gen neut hybridize inf gen) N fem 12 * :according
Pre gen

(3.5)

(der mas percentage mas der gen fem plur mature pastpt gen fem plur capsule
gen fem plur) N mas 12 *
-(der fem number fem der gen plur mas grain gen neut plur der gen fem plur
seed gen fem plur
one gen fem capsule gen fem) N fem 12 *
.different Aj *
,(der fem time fem der gen neut hybridize inf gen) N fem 12 * :according
Pre gen

(3.6)

The numbers following the N word classes indicate the number of potentially inflectable words in the preceding phrase.

The two following stages deal, as before, with prepositional phrases and coordinate nouns:-

(der mas percentage mas der gen fem plur mature pastpt gen fem plur capsule
gen fem plur) N mas 12 *
-(der fem number fem der gen plur mas grain gen neut plur der gen fem plur
seed gen fem plur
one gen fem capsule gen fem) N fem 12 *
.different Aj *
.according der gen fem time gen fem der gen neut hybridize inf gen) Av

(3.7)

(der mas percentage mas der gen fem plur gen mature pastpt gen fem plur
capsule gen fem plur
und der fem number fem der gen plur mas grain gen neut plur der gen fem plur
seed gen fem plur
one gen fem capsule gen fem) N plur 12 und 12 *
.different Aj *
.according der gen fem time gen fem der gen neut hybridize inf gen) Av

(3.8)

Having reduced to the word class sequence N plur.Aj Av, the final reduction can be made:-

(der mas percentage mas der gen fem plur gen mature pastpt gen fem plur capsule
gen fem plur
und der fem number fem der gen plur mas grain gen neut plur der gen fem plur
seed gen fem plur
one gen fem capsule gen fem
sein plur
according der gen fem time gen fem der gen neut
hybridize inf gen
different)

(3.9)

The chunk-category interactions, though numerous, present no difficulty, and the final output is:-

der Prozentsatz der ge-reif-ten Kapsel -n
und die Zahl der Gran-e der Same -n
ein-er Kapsel -n
sind
gemäss der Zeit des Bastardier-ens
verschieden

(3.10)

Translation from Nude to Latin

The first stage provides:-

ratio per centum nz'P"PlOO Nlab fem III * .,capsul Xp"CL'HI'Q N fem I * :matur
B"f'CL V I *
-numer P N ab mas II * .gran Nz'cN neut II * .semin Xp'CL N neut III *
.capsul Xp"CL'HI'Q N fem I * .un Pz Aj sing irr * .divers Sz Aj I *
,tempor T N ab neut III * ,:hybridiz CL'Sz V I * :secundum C3 Pre acc

(4.1)

or

percentage N1 ab fem III * .,capsule N fem I * :mature V I *
number N ab mas II * .grain N neut II * .seed N neut III *
.capsule N fem I * .one Aj sing irr
.different Aj I *
,time N ab neut III * ,:hybridize V I * :according Pre acc

(4.2)

The stages in reduction are very similar to those already described. The definite article does not have to be inserted:-

percentage N1 ab fem II * .,capsule N fem I * :mature V I *
number N ab mas II * .grain N neut II * .seed N neut III *
.capsule fem I one fem irr) N2 fem *
.different Aj I *
,time N ab neut III * ,:hybridize V I * :according Pre acc

(4.3)

percentage N1 ab fem III * .(capsule acc fem plur I mature ger I) N2 *
-(number mas II grain gen neut plur II) N1 mas * seed N neut III *
.capsule fem I one fem irr) N2 fem *
.different Aj I *
,(time neut III hybridize ger gen I) N1 neut * :according Pre acc

(4.4)

(percentage fem III capsule acc fem plur I mature ger gen I) N1 fem *
-(number mas II grain gen neut plur II seed gen neut plur III) N1 mas *
.capsule gen fem I one gen fem irr) N1 mas *
.different Aj I *
,(time neut III hybridize ger gen I) N1 neut * :according Pre acc

(4.5)

(percentage fem III capsule acc fem plur I mature ger gen I) N1 fem *
-(number mas II grain gen neut plur II seed gen neut plur III)
.capsule gen fem I one gen fem irr) N1 mas *
.different Aj I *
,(according time acc neut III hybridize ger gen I) Av

(4.6)

(percentage fem III capsule acc fem plur I mature ger gen I
et number mas II grain gen neut plur II seed gen neut plur III
.capsule gen fem I one gen fem irr) N plur I fem et I mas *
.different Aj I *
,(according time acc neut III hybridize ger gen I) Av

(4.8)

(percentage fem III capsule acc fem plur I mature ger gen I
et number mas II grain gen neut plur II seed gen neut plur III
capsule gen fem I one gen fem irr
according time acc neut III hybridize ger gen I
different neut plur I esse plur)

(4.9)

After matching against the Nude-Latin dictionary for the chunk-category interactions the output is:-

ratio per centum capsul -as matur -andi
et numer -us gran -orum semin -um
capsul -ae unius
secundum tempor -em hybridiz -andi
divers -a sunt

Translation from Nude to Welsh

The preliminary matching against the Nude-Welsh dictionary gives:-

anran nz'P"P100 N ab fem c * .,asgl Xp"CL'HI'Q N mas m * :aedd fed B"f'CL V u *
-if P N ab mas rh * .ronyn Mz'cN mas g nau * had Xp'CL N mas *
.asgl Xp"CL'HI'Q N mas m * un Pz Aj sing *
.wahanol Sz Aj g *
,amser T N ab mas oedd * ,:roes c i rhywiau Vl * :yn ôl C3 Pre

(5.1)

or

percentage N ab fem c * .,capsule N mas m * :mature V u *
-number N ab mas rh * .grain N mas g nau * seed N mas *
.capsule N mas m * one Aj sing *
.different Aj g *
,time N ab mas oedd * ,:hybridize Vl c i * :according Pre

(5.2)

The stages of reduction are fundamentally similar to the preceding cases with the addition of instructions for initial mutation:-

percentage N ab fem c * .,capsule N mas m * :mature V u *
-number N ab mas rh * .grain N mas g nau * seed N mas *
.(one mas capsule mas m) N12 fem *
.different Aj g *
,time N ab mas oedd * ,:hybridize Vl c i * :according Pre

(5.3)

percentage N ab fem c * .(mature inf u capsule mas plur m) N2 fem plur *
-(number mas rh grain plur g nau seed plur) N1 mas *
.(one mas capsule mas m) N12 fem *
.different Aj g *
,(time mas oedd hybridize inf c i) :according Pre

(5.4)

(percentage fem c mature inf u capsule mas plur m) N1 fem *
-(number mas rh grain plur g nau seed plur) N1 mas *
.(one mas capsule mas m) N12 fem *
.different Aj g *
,(time mas oedd hybridize inf c i) :according Pre

(5.5)

(percentage fem c mature inf u capsule mas plur m) N1 fem *
-(number mas rh grain plur g nau seed plur
one mas capsule mas m) N mas l *
.different Aj g *
,(time mas oedd hybridize inf c i) :according Pre

(5.6)

(percentage fem c mature inf u capsule mas plur m) N1 fem *
-(number mas rh grain plur g nau seed plur
one mas capsule mas m) N mas l *
.different Aj g *
,(according time mas oedd hybridize inf c i) Av

(5.7)

(percentage fem c mature inf u capsule mas plur m
a number mas rh asp grain plur g nau seed plur
one mas capsule mas m) N plur l fem a l mas
.different Aj g *

.(according time mas oedd hybridize inf c i) Av

(5.8)

(y mae
percentage fem c mature inf u capsule mas plur m
a number mas rh asp grain plur g nau seed plur
one mas capsule mas m
yn different g soft
according time mas oedd hybridize inf c i)

(5.9)

After ascertaining the chunk-category interactions, the final output is

y mae
c- anran aeddfed -u m- asgl -au
a rh- if gronyn -nau had -au
un m- asgl
yn wahanol
yn ôl amser c- roes -i rhywiau

(5.10)

List of Nude elements

B	becoming, change	p	plant
c	straight, plane	P	plurality, group, number
C	causation, influence	Q	hard, firm
f	possibility, potentiality	S	same, equal
H	pertain	T	time, period, duration
I	in, inside	u	elongate
L	living, alive	x	textile
M	much, more, great	X	part, component
n	near, adjacent, together	z	negation, opposite, contrary
N	contact, adhere, attach		

List of category indications

ab	abstract	N	noun
acc	accusative	nau	plural suffix (Welsh)
Aj	adjective	neut	neuter
asp	aspirate initial mutation (Welsh)	Num	numeral
C	coordinate conjunction	oedd	plural suffix (Welsh)
c	c-g-ngh-ch initial series (Welsh)	Pastpt	past participle
e	verb suffix (English)	plur	plural
fem	feminine	Post	post position
g	g-ng initial series (Welsh)	Pre	preposition
gen	genitive	prept	present participle
ger	gerund	rh	rh-r initial series (Welsh)
i	infinitive suffix (Welsh)	ru	verb suffix (Japanese)
inf	infinitive	sing	singularizer
irr	irregular	soft	soft initial mutation (Welsh)
m	m-f initial series (Welsh)	u	infinitive suffix (Welsh)
mas	masculine	Vt	verb transforming suffix (Japanese)

Arabic numbers indicate the position of inflectable chunks. Roman numbers indicate morphological subcategories, e.g. declensions.

Summary

A general programme for mechanical translation via an algebraic interlingua called Nude is described.

From a Japanese passage conventionally transliterated as: KETSU SAKU HO CO HEI ni ICHI SAKU To² ri SHU SHI RIU SU² ha KO HAI JI KI ni yo tsu te I ru, the following specimen translations into English, German, Latin and Welsh, respectively, were obtained:

- 10 -
(1) the percentage of matured capsules and the number of grains of seeds of one capsule are different according to the time of hybridizing.

(2) der Prozentsatz der gereiften Kapseln und die Zahl der Grane der Samen einer Kapseln sind gemäss der Zeit des Bastardierens verschieden.

(3) ratio per centum capsulas maturandi et numerus granorum seminum capsulae unius secundum temporem hybridizandi diversa sunt.

(4) y mae canran oedd fedu masglau a rhif gronynnau hadau un masgl yn wahanol yn ôl amser croesi rhywiau.

The method outlined is illustrative only and would require intensive development for general application. It seems however that interlingual translation has great potentialities and should be seriously considered in any mechanical translation research programme.

Grateful acknowledgement is made to Mr. M. Rees for assistance with the translation into Welsh.

References

Richens, R.H. and Booth, A.D., 1955, Some methods of mechanized translation.
Machine Translation of Languages, edited by W.N. Locke and A.D. Booth

Richens, R.H. (in the press) Preprogramming for mechanical translation.