MACHINE TRANSLATION REVIEW

The *Machine Translation Review* incorporates the Newsletter of the Natural Language Translation Specialist Group of the British Computer Society and appears annually in December.

The Review welcomes contributions, articles, book reviews, advertisements, and all items of information relating to the processing and translation of natural language. Contributions and correspondence should be addressed to:

Derek Lewis
The Editor
Machine Translation Review
School of Modern Languages
Queen's Building
University of Exeter
Exeter
EX4 4QH
United Kingdom

Tel: +44 (0)1392 264325
E-mail: D.R.Lewis@exeter.ac.uk

Please note: From April 2000 this Review will be published electronically and will be available on our web site at the British Computer Society (see page 5). The format will be in HTML, in the same way as some back copies have already been stored electronically, so it will be easy for readers to print copies if they wish. Each section will be separate so readers may print selected parts only.

Some copies will be printed for the Copyright libraries and for purchase at a modest price plus postage and packing for those without electronic access. Members of the Natural Language Translation Specialist Group of the British Computer Society will be advised of each issue.

Contents

Group News and Information

Letter from the Chairman

The British Computer Society
Charity No. 292786

THE NATURAL LANGUAGE TRANSLATION SPECIALIST GROUP

Please reply to:
72 Brattle Wood
Sevenoaks
Kent, TN13 1QU
Tel: 01732 455446
E-mail: wiggjd@bcs.org.uk

We have enjoyed another successful year culminating with the Workshop on Teaching MT organised jointly with the European Association for MachineTranslation (EAMT) which was well attended in November. The Proceedings will soon be available from the EAMT website at http://www.eamt.org.

I have been experimenting with some of the free translation systems which have been available on the internet for a while now. I found them interesting, entertaining, educational, and surprisingly enough, sometimes useful. We are now considering a project to find out the best way of using them. Anybody interested in this research should contact me as soon as possible. We hope to publish the results on our website in due course.

I would like to take this opportunity of reminding you that our web master, Roger Harris, would very much welcome more information or updates to be loaded onto our website. If you looked there for something but did not find it but you found it somewhere else please let him know so it could be included in future for other people.

I have just realised that I gave an incorrect address for our e-mail list in the last issue. It should have been bcs-mt@jiscmail.ac.uk. Perhaps that accounts for the list being rather quiet last year. To join, send an e-mail to jiscmail@jiscmail.ac.uk with 'join bcs-mt your_firstname your_surname' in the message area. Leave the subject line blank. Please contribute to the list via bcs-mt@jiscmail.ac.uk. Subscription to the list is free but don't forget that postings are automatically circulated to all subscribers.

Once again I must ask you to consider contributing to this Review. We would still welcome more articles, papers and reports on the theory and practice of machine translation and related subjects such as computer assisted language teaching, computer based dictionaries and aspects of multi-linguality in computing, etc. We welcome papers from staff and students in linguistics and related disciplines, and from translators and any other users of MT software.

Members do not need to live near London to assist the Committee. We do not always have sufficient funds to pay travel expenses for all Committee members to attend meetings, but we still welcome Correspondent members. Correspondent committee members are otherwise treated as full members of the committee and kept advised of all committee business. Anyone interested in helping should contact me or any other Committee member.

David Wigg, January 2003

The Committee

The telephone numbers and e-mail addresses of the Committee are as follows:

David Wigg (Chair & Treasurer)	Tel.: +44 (0)1732 455446 (H)
	Tel.: +44 (0)207 815 7472 (W)
	E-mail: wiggjd@bcs.org.uk
Monique L'Huillier (Secretary)	Tel.: +44 (0)1276 20488 (H)
	Tel.: +44 (0)1784 443243 (W)
	E-mail: m.l'huillier@rhul.ac.uk
Derek Lewis (Editor)	Tel.: +44 (0)1404 814186 (H)
	Tel.: +44 (0)1392 264325 (W)
	E-mail: d.r.lewis@exeter.ac.uk
Roger Harris (Webmaster)	Tel.: +44 (0)208 800 2903 (H)
	E-mail: rwsh@bcs.org.uk
Douglas Clarke	Tel.: +44 (0)1908 373141
Ian Kelly	Tel.: +44 (0)1276 857599
	E-mail: idkk@idkk.com
Veronica Lawson	Tel.: +44 (0)207 7359060
	E-mail: veronica_1@compuserve.com
Correspondent Members:	
Gareth Evans (Minority Languages)	E-mail: gevans@ubs-wsc.org
Ruslan Mitkov	Tel: +44 (0)1902 322471 (W)
	E-mail: r.mitkov@wlv.ac.uk
Heather Fulford	E-mail: h.fulford@lboro.ac.uk
Mark Stevenson	E-mail: m.stevenson@dcs.shef.ac.uk

BCS Library

Books kindly donated by members are passed to the BCS library at the IEE, Savoy Place, London, WC2R 0BL, UK (tel: +44 (0)207 240 1871; fax: +44 (0)207 497 3557). Members of the BCS may borrow books from this library either in person or by post. All they have to provide is their membership number. The library is open Monday to Friday, 9.00 am - 5.00 pm.

Website

The website address of the BCS-NLTSG is: http://www.bcs.org.uk/siggroup/nalatran

The Pattern-Based French-to-Polish Machine Translation System

by

Dr Miroslaw Gajer

Technical University in Cracow
Department of Control Science
al. Mickiewicza 30
30-059 Kraków
Poland
mgajer@ia.agh.edu.pl

Abstract

The field of machine translation now has over 50 years of history, and despite great scientific effort is still far a final and fully satisfactory solution. Several years ago a case-based machine translation technique was proposed, and proved to be a serious alternative to existing machine translation methods. However, case-based machine translation has never been implemented for the Polish language. This author has investigated the possibilities of implementing the case-based machine translation technique for the Polish languages, leading him to the conclusion that this technique must be modified thoroughly if it is to add to the effective construction of a machine translation system for unrestricted text. The modification of case-based machine translation technique for the Polish language led finally to a translation technique, which the author has named pattern-based machine translation. This paper demonstrates the implementation of pattern-based machine translation technique for French to Polish translation.

Introduction

Machine translation is a science about writing computer programs that are able to translate in a fully automatic way between natural languages, e.g. between Spanish and Korean [1]. It was suggested by an American, Warren Weaver, who on 15th July 1949 sent his famous memorandum to the Rockefeller Foundation that started scientific research in the field [2]. The first public demonstration of an operating machine translation system took place in New York on 7th December 1954. During this demonstration the machine translation system translated 49 simple sentences from Russian to English.

But after over 50 years of intensive scientific research machine translation is still far from a satisfactory solution [4]. Several years ago case-based or example-based machine translation techniques emerged as a very serious alternative to other methods used so far for automating the translation process [3], [5], [6], [7].

The main idea of CBMT (Case-Base Machine Translation) is as follows. Let there be given any bilingual text, for example the following English and Polish text:

There are certain themes of which the interest is all-absorbing, but which are too entirely horrible for the purposes of legitimate fiction. These the mere romanticist must eschew, if he do not wish to offend, or to disgust. They are with propriety handled only when the severity and majesty of truth sanctify and sustain them.

and

Istnieją pewne tematy przejmujące do głębi, wszelako nazbyt okropne, izby mogły stać się przedmiotem rzetelnej literatury powieściowej. Rzetelni pisarze muszą ich unikać, jeśli nie chcą wywołać zgorszenia lub niesmaku. Godzi się je podejmować tylko w takich razach, gdy uświęca je i usprawiedliwia surowe dostojeństwo prawdy.

The first text is the beginning of the story by Edgar Allan Poe entitled 'The Premature Burial' and the second one is its Polish translation made by a prominent translator, Stanisław Wyrzykowski.

In this bilingual text the English phrase 'they are with proprietary handled' was translated into Polish by a human professional translator as 'godzi się je podejmować'. So the idea of the CBMT is such that, while translating any other English text into Polish, if the English phrase 'they are with proprietary handled' appears again anywhere else, it will be substituted the same way as it had been done before by a human translator, i.e. it will be substituted by 'godzi się je podejmować' . The translation obtained in such a manner should still be correct.

Of course, there is no certainty that the translation obtained by the use of CBMT will always and under any circumstances be correct. The main reason for this is natural language ambiguity, which is a permanent feature of any human language. Even translations of whole sentences can be totally different, depending on the context in which these sentences are used. A good example for this is the Dutch sentence:

Als ik in Amsterdam werkte kopte ik en auto.

This sentence can be translated into English as:

If I worked in Amsterdam I would buy a car.

or as:

When I worked in Amsterdam I bought a car.

Everything depends on the context in which the sentence is used. So we can see that the CBMT is not always able to return a correct translation, but the probability that the translation will be correct is still very high, and generally the quality of a text translated with the use of the CBMT is much higher than one obtained with the use of other, e.g. transfer or interlingua, methods.

The Implementation of CBMT for the Polish Language

The Polish language, which belongs to the group of Slavonic languages, posseses a quite complex grammar. The main feature of Polish is its inflexional nature, which makes learning it extremely difficult for a foreigner. In Polish grammar there can be found such phenomena as declination of nouns (there are 6 cases) and conjugation of verbs. Also there are singular and plural forms of words and 3 genders (masculine, feminine and neutral).

These phenomena make the direct usage of CBMT for the Polish language difficult and non-effective. In order to implement CBMT for Polish it must be thoroughly modified. This author has proposed such a modification, which takes into account the peculiarities of the Polish grammar and allows for its effective usage for the purpose of translating into Polish from most West European languages.

In the system developed by this author there exist 3 different types of translation examples, which are further called translation patterns. The way in which these patterns are translated into Polish depends on the actual value of 3 attributes: <case>, <number>, and <gender>.

The first type of translation pattern is the so called declination-type translation pattern, which has the following form:

'source language phrase'	< case >
'translation into target language for <case>=1'	1
'translation into target language for <case>=2'	2
'translation into target language for <case>=3'	3
'translation into target language for <case>=4'	4
'translation into target language for <case>=5'	5
'translation into target language for <case>=6'	6
< case > = x; < number > = y; < gender > = z;	

The second type of translation pattern is the so called conjugation-type translation pattern:

'source language text'	< number >	< gender >
'translation into target language for <number>=1 and <gender>=1'	1	1
'translation into target language for <number>=1 and <gender>=2'	1	2
'translation into target language for <number>=1 and <gender>=3'	1	3
'translation into target language for <number>=2 and <gender>=1'	2	1
'translation into target language for <number>=2 and <gender>=2'	2	2
'translation into target language for <number>=2 and <gender>=3'	2	3
< case > = x; < number > = y; < gender > = z;		

The third type of translation pattern is the non-flexion-type translation pattern:

'source language text'
'translation into target language'
< case > = x; < number > = y; < gender > = z;

As we can see, each of the translation patterns has in its bottom row the optional information about the values that must be assigned to the <case>, <number>, and <gender> attributes, so that the grammatical relations between subject, verb, and object should be preserved. This part of the translation pattern is optional, i.e. not all the attributes must be assigned values in every translation pattern. In some special cases no attributes can be assigned values at all. The <case> attribute is assigned values from 1 to 6 (depending on the grammatical case). The attribute <number> is assigned value 1 for singular or 2 for plural. And the attribute <gender> is assigned value 1 for masculine, 2 for feminine, or 3 for neutral.

The pattern-based machine translation technique was implemented by this author for the system that translates from French to Polish. The manner in which the system operates will be described thoroughly in the next paragraph.

The French-to-Polish Machine Translation System in Operation

In order to illustrate how the pattern-based French-to-Polish machine translation system operates, the following quite complicated French text will be considered for translation:

Chopin naquit en 1810 à Zelazowa Wola près de Varsovie. Son père y enseignait le français dans une famille riche. Ensuite, sa famille vint s'installer à Varsovie. Il étudia le piano et suivit les cours du Conservatoire. Aussitôt qu'il eut commencé ses études, ses maîtres s'aperçurent qu'il avait du génie. Lorsqu'il eut fini ses études au Conservatoire de Varsovie, il quitta la Pologne et il vint à Paris où il resta jusqu'à la fin de sa vie. Il y donna d'abord des leçons particulières de piano et y il eut des jours difficiles. Mais, dès qu'il eut donné ses premiers concerts, il devint célèbre. Il fit la connaissance de la romancière George Sand avec laquelle il resta en amitié de longues années. Ils voyagèrent ensemble à Majorque. L'état de santé de Chopin s'aggrava, dès qu'ils furent rentrés en France. La maladie de poitrine dont il souffrait empirait de jour en jour. Malgré ses souffrances, il continua à composer les chefs-d'oevre qui l'ont rendu immortel. Il mourut à Paris en 1849, en pensant à sa patrie, la Pologne.

Les Polonais ont depuis longtemps participé à la vie française. Et le souvenir de certains d'entre eux reste toujours vivant dans la mémoir des Français. En effet, les Français connaissent les noms des Polonais illustre comme Thadée Kościuszko, Frédéric Chopin et Marie Curie-Sklodowska. Ce sont là des Polonais qui se sont engagés aux côtés de la France populaire. Mais pour les Français, c'est Adam Mickiewicz qui est l'incarnation vivante de son pays. Victor Hugo a dit que parle de Mickiewicz, c'est parle du beau, du juste et du vrai: c'est parle du droit, dont il a été le soldat, du devoir, dont il a été le héros, de la liberté, dont il a été l'apôtre, et de la délivrance, dont il a été le précurseur. Victor Hugo et Adam Mickiewicz symboliset bien l'amitié entre nos deux peuples.

The machine translation system looks up in the database the following translation patterns and substitutes them with their Polish equivalents, as follows:

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

1) the declination-type translation pattern

Chopin	< case >
Chopin	1
Chopina	2
Chopinowi	3
Chopina	4
Chopinem	5
Chopinie	6
< number > = 1;	
< gender > = 1;	

2) the conjugation-type translation pattern

naquit	< number >	< gender >
urodził się	1	1

urodziła się	1	2
urodziło się	1	3
urodzili się	2	1
urodziły się	2	2
urodziły się	2	3

3) the non-flexion-type translation pattern

| en 1810 |
| w roku 1810 |

4) the non-flexion-type translation pattern

| à Żelazowa Wola |
| w Żelazowej Woli |

5) the non-flexion-type translation pattern

| près de |
| blisko |
| < case > = 2; |

6) the declination-type translation pattern

Varsovie	< case >
Warszawa	1
Warszawy	2
Warszawie	3
Warszawę	4
Warszawą	5
Warszawie	6
< number > = 1; < gender > = 2;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

| < case > = 1; |

7) the declination-type translation pattern

son père	< case >
jego ojciec	1
jego ojca	2
jego ojcu	3
jego ojca	4
jego ojcem	5
jego ojcu	6
< number > = 1; < gender > = 1;	

8) the conjugation-type translation pattern

y enseignait	< number >	< gender >
uczył tam	1	1
uczyła tam	1	2
uczyło tam	1	3
uczyli tam	2	1
uczyły tam	2	2
uczyły tam	2	3
< case > = 2;		

9) the declination-type translation pattern

le français	< case >
język francuski	1
języka francuskiego	2
językowi francuskiemu	3
język francuski	4
językiem francuskim	5
języku francuskims	6
< number > = 1;	
< gender > = 1;	

10) the non-flexion-type translation pattern

dans
w
< case > = 6;

11) the declination-type translation pattern

une famille riche	< case >
bogata rodzina	1
bogatej rodziny	2
bogatej rodzinie	3
bogatą rodzinę	4
bogatą rodziną	5
bogatej rodzinie	6
< number > = 1;	
< gender > = 2;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

12) the non-flexion-type translation pattern

ensuite
następnie

13) the declination-type translation pattern

sa famille	< case >
jego rodzina	1
jego rodziny	2
jego rodzinie	3

jego rodzinę	4
jego rodziną	5
jego rodzinie	6
< number > = 1; < gender > = 2;	

14) the conjugation-type translation pattern

vint s'installer à	< number >	< gender >
przeniósł się do	1	1
przeniosła się do	1	2
przeniosło się do	1	3
przenieśli się do	2	1
przeniosły się do	2	2
przeniosły się do	2	3
< case > = 2;		

15) the declination-type translation pattern

Varsovie	< case >
Warszawa	1
Warszawy	2
Warszawie	3
Warszawę	4
Warszawą	5
Warszawie	6
< number > = 1; < gender > = 2;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

16) the non-flexion-type translation pattern

il
on
< number > = 1; < gender > = 1;

17) the conjugation-type translation pattern

étudia le piano	< number >	< gender >
uczył się gry na fortepianie	1	1
uczył asie gry na fortepianie	1	2
uczyło się gry na fortepianie	1	3
uczyli się gry na fortepianie	2	1
uczyły się gry na fortepianie	2	2
uczyły się gry na fortepianie	2	3

18) the non-flexion-type translation pattern

et
i

19) the conjugation-type translation pattern

suivit les cours du	< number >	< gender >
uczęszczał na zajęcia w	1	1
uczęszczała na zajęcia w	1	2
uczęszczało na zajęcia w	1	3
uczęszczali na zajęcia w	2	1
uczęszczały na zajęcia w	2	2
uczęszczały na zajęcia w	2	3
< case > = 6;		

20) the declination-type translation pattern

Conservatoire	< case >
Konserwatorium	1
Konserwatorium	2
Konserwatorium	3
Konserwatorium	4
Konserwatorium	5
Konserwatorium	6
< number > = 1; < gender > = 3;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

21) the non-flexion-type translation pattern

aussitôt
skoro tylko

22) the non-flexion-type translation pattern

qu'il eut commencé ses études
rozpoczął studia

23) the declination-type translation pattern

ses maîtres	< case >
jego nauczyciele	1
jego nauczycieli	2
jego nauczycielom	3
jego nauczyciele	4
jego nauczycielami	5
jego nauczycielach	6
< number > = 2; < gender > = 1;	

24) the conjugation-type translation pattern

s'aperçurent	< number >	< gender >
zauważył	1	1
zauważyła	1	2
zauważyło	1	3

zauważyli	2	1
zauważyły	2	2
zauważyły	2	3

25) the non-flexion-type translation pattern

| qu'il avait du génie |
| że był genialnie uzdolniony |

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

| < case > = 1; |

26) the non-flexion-type translation pattern

| lorsqu'il eut fini |
| kiedy ukończył |
| < case > = 4; |

27) the declination-type translation pattern

ses études	< case >
swoje studia	1
swoich studiów	2
swoim studiom	3
swoje studia	4
swoimi studiami	5
swoich studiach	6
< number > = 2; < gender > = 3;	

28) the non-flexion-type translation pattern

| au Conservatoire de Varsovie |
| w Konserwatorium Warszawskim |

29) the non-flexion-type translation pattern

| il |
| on |
| < number > = 1; |
| < gender > = 1; |

30) the conjugation-type translation pattern

quitta	< number >	< gender >
opuścił	1	1
opuściła	1	2
opuściło	1	3
opuścili	2	1
opuściły	2	2
opuściły	2	3
< case > = 4;		

31) the declination-type translation pattern

la Pologne	< case >
Polska	1
Polski	2
Polsce	3
Polskę	4
Polską	5
Polsce	6
< number > = 1;	
< gender > = 2;	

32) the non-flexion-type translation pattern

et
i

33) the non-flexion-type translation pattern

il
on
< number > = 1;
< gender > = 1;

34) the conjugation-type translation pattern

vint à	< number >	< gender >
wyjechał do	1	1
wyjechała do	1	2
wyjechało do	1	3
wyjechali do	2	1
wyjechały do	2	2
wyjechały do	2	3
< case > = 2;		

35) the declination-type translation pattern

Paris	< case >
Paryż	1
Paryża	2
Paryżowi	3
Paryż	4
Paryżem	5
Paryżu	6
< number > = 1;	
< gender > = 1;	

36) the non-flexion-type translation pattern

où
gdzie

37) the non-flexion-type translation pattern

il	
on	
< number > = 1;	
< gender > = 1;	

38) the conjugation-type translation pattern

resta	< number >	< gender >
pozostał	1	1
pozostała	1	2
pozostało	1	3
pozostali	2	1
pozostały	2	2
pozostały	2	3

39) the non-flexion-type translation pattern

jusqu'à
do
< case > = 2;

40) the declination-type translation pattern

la fin	< case >
koniec	1
końca	2
końcowi	3
koniec	4
końcem	5
końcu	6
< number > = 1;	
< gender > = 1;	

41) the non-flexion-type translation pattern

de sa vie
swojego życia

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

42) the non-flexion-type translation pattern

il
on
< number > = 1;
< gender > = 1;

43) the conjugation-type translation pattern

y donna	< number >	< gender >
dawał tam	1	1
dawała tam	1	2
dawało tam	1	3
dawali tam	2	1

dawały tam	2	2
dawały tam	2	3
< case > = 4;		

44) the non-flexion-type translation pattern

| d'abord |
| najpierw |

45) the declination-type translation pattern

des leçons particulières de piano	< case >
prywatne lekcje gry na fortepianie	1
prywatnych lekcji gry na fortepianie	2
prywatnym lekcjom gry na fortepianie	3
prywatne lekcje gry na fortepianie	4
prywatnymi lekcjami gry na fortepianie	5
prywatnych lekcjach gry na fortepianie	6
< number > = 2;	
< gender > = 2;	

46) the non-flexion-type translation pattern

| et |
| i |

47) the non-flexion-type translation pattern

| y il eut des jours difficiles |
| przeżywał tam ciężkie dni |

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

| < case > = 1; |

48) the non-flexion-type translation pattern

| mais |
| ale |

49) the non-flexion-type translation pattern

| dès qu'il eut donné |
| od chwili gdy dał |
| < case > = 4; |

50) the declination-type translation pattern

ses premiers concerts	< case >
swoje pierwsze koncerty	1
swoich pierwszych koncertów	2
swoim pierwszym koncertom	3
swoje pierwsze koncerty	4
swoimi pierwszymi koncertami	5
swoich pierwszych koncertach	6
< number > = 2;	

17

< gender > = 3;

51) the non-flexion-type translation pattern

il
on
< number > = 1;
< gender > = 1;

52) the conjugation-type translation pattern

devint célèbre	< number >	< gender >
stał się sławny	1	1
stała się sławna	1	2
stało się sławne	1	3
stali się sławni	2	1
stały się sławne	2	2
stały się sławne	2	3

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

53) the non-flexion-type translation pattern

il
on
< number > = 1;
< gender > = 1;

54) the conjugation-type translation pattern

fit la connaissance de	< number >	< gender >
zawarł znajomość z	1	1
zawarła znajomość z	1	2
zawarło znajomość z	1	3
zawarli znajomość z	2	1
zawarły znajomość z	2	2
zawarły znajomość z	2	3
< case > = 5;		

55) the declination-type translation pattern

la romancière George Sand	< case >
powieściopisarka George Sand	1
powieściopisarki George Sand	2
powieściopisarce George Sand	3
powieściopisarkę George Sand	4
powieściopisarką George Sand	5
powieściopisarce George Sand	6
< number > = 1;	
< gender > = 2;	

56) the non-flexion-type translation pattern

avec laquelle
z którą

57) the non-flexion-type translation pattern

il
on
< number > = 1;
< gender > = 1;

58) the conjugation-type translation pattern

resta en amitié	< number >	< gender >
pozostał w przyjaźni	1	1
pozostała w przyjaźni	1	2
pozostało w przyjaźni	1	3
pozostali w przyjaźni	2	1
pozostały w przyjaźni	2	2
pozostały w przyjaźni	2	3

59) the non-flexion-type translation pattern

de longues années
przez długie lata

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

60) the non-flexion-type translation pattern

ils
oni
< number > = 2;
< gender > = 1;

61) the conjugation-type translation pattern

voyagèrent	< number >	< gender >
podróżował	1	1
podróżowała	1	2
podróżowało	1	3
podróżowali	2	1
podróżowały	2	2
podróżowały	2	3

62) the non-flexion-type translation pattern

ensamble
razem

63) the non-flexion-type translation pattern

à Majorque
na Majorkę

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

64) the declination-type translation pattern

l'état de santé	< case >
stan zdrowia	1
stanu zdrowia	2
stanowi zdrowia	3
stan zdrowia	4
stanem zdrowia	5
stanie zdrowia	6
< number > = 1; < gender > = 1;	

65) the non-flexion-type translation pattern

de Chopin
Chopina

66) the conjugation-type translation pattern

s'aggrava	< number >	< gender >
pogorszył się	1	1
pogorszyła się	1	2
pogorszyło się	1	3
pogorszyli się	2	1
pogorszyły się	2	2
pogorszyły się	2	3

67) the non-flexion-type translation pattern

dès qu'ils furent rentrés en
po tym jak wrócił do
< case > = 2;

68) the declination-type translation pattern

France	< case >
Francja	1
Francji	2
Francji	3
Francję	4
Francją	5
Francji	6
< number > = 1; < gender > = 2;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

69) the declination-type translation pattern

la maladie de poitrine	< case >
choroba płuc	1
choroby płuc	2
chorobie płuc	3
chorobę płuc	4
chorobą płuc	5
chorobie płuc	6
< number > = 1; < gender > = 2;	

70) the conjugation-type translation pattern

dont il souffrait	< number >	< gender >
na który cierpiał	1	1
na którą cierpiał	1	2
na które cierpiał	1	3
na których cierpiał	2	1
na które cierpiał	2	2
na które cierpiał	2	3

71) the conjugation-type translation pattern

empirait	< number >	< gender >
pogłębiał się	1	1
pogłębiała się	1	2
pogłębiało się	1	3
pogłębiali się	2	1
pogłębiały się	2	2
pogłębiały się	2	3

72) the non-flexion-type translation pattern

de jour en jour
z dnia na dzień

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

73) the non-flexion-type translation pattern

malgré
pomimo
< case > = 2;

74) the declination-type translation pattern

ses souffrances	< case >
jego cierpienia	1
jego cierpień	2
jego cierpieniom	3
jego cierpienia	4
jego cierpieniami	5
jego cierpieniach	6
< number > = 2; < gender > = 3;	

75) the non-flexion-type translation pattern

il
on
< number > = 1;
< gender > = 1;

76) the conjugation-type translation pattern

continua à composer	< number >	< gender >
nadal komponował	1	1
nadal komponowała	1	2
nadal komponowało	1	3
nadal komponowali	2	1
nadal komponowały	2	2
nadal komponowały	2	3
< case > = 4;		

77) the declination-type translation pattern

les chefs-d'oevre	< case >
arcydzieła	1
arcydzieł	2
arcydziełom	3
arcydzieła	4
arcydziełami	5
arcydziełach	6
< number > = 2;	
< gender > = 3;	

78) the conjugation-type translation pattern

qui l'ont rendu immortel	< number >	< gender >
który uczynił go nieśmiertelnym	1	1
która uczyniła go nieśmiertelnym	1	2
które uczyniło go nieśmiertelnym	1	3
którzy uczynili go nieśmiertelnym	2	1
które uczyniły go nieśmiertelnym	2	2
które uczyniły go nieśmiertelnym	2	3

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

79) the non-flexion-type translation pattern

il
on
< number > = 1;
< gender > = 1;

80) the conjugation-type translation pattern

mourut à	< number >	< gender >
umarł w	1	1
umarła w	1	2
umarło w	1	3
umarli w	2	1
umarły w	2	2
umarły w	2	3
< case > = 6;		

81) the declination-type translation pattern

Paris	< case >
Paryż	1
Paryża	2
Paryżowi	3
Paryż	4
Paryżem	5
Paryżu	6
< number > = 1;	
< gender > = 1;	

82) the non-flexion-type translation pattern

en 1849
w roku 1849

83) the non-flexion-type translation pattern

en pensant à
myśląc o
< case > = 6;

84) the declination-type translation pattern

sa patrie	< case >
swoja ojczyzna	1
swojej ojczyzny	2
swojej ojczyźnie	3
swoją ojczyznę	4
swoja ojczyzną	5
swojej ojczyźnie	6
< number > = 1;	
< gender > = 2;	

85) the declination-type translation pattern

la Pologne	< case >
Polska	1
polski	2
Polsce	3
Polskę	4
Polską	5
Polsce	6
< number > = 1;	
< gender > = 2;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

86) the declination-type translation pattern

les Polonais	< case >
Polacy	1
Polaków	2
Polakom	3
Polaków	4
Polakami	5
Polakach	6
< number > = 2; < gender > = 1;	

87) the conjugation-type translation pattern

ont depuis longtemps participé à	< number >	< gender >
od dawna uczestniczył w	1	1
od dawna uczestniczyła w	1	2
od dawna uczestniczyło w	1	3
od dawna uczestniczyli w	2	1
od dawna uczestniczyły w	2	2
od dawna uczestniczyły w	2	3
< case > = 6;		

88) the declination-type translation pattern

la vie française	< case >
życie Francji	1
życia Francji	2
życiu Francji	3
życie Francji	4
życiem Francji	5
życiu Francji	6
< number > = 1; < gender > = 3;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

89) the non-flexion-type translation pattern

et
i

90) the declination-type translation pattern

le souvenir de certains d'entre eux	< case >
pamiątki po niektórych z nich	1
pamiątek po niektórych z nich	2
pamiątkom po niektórych z nich	3
pamiątki po niektórych z nich	4
pamiątkami po niektórych z nich	5

pamiątkach po niektórych z nich	6
< number > = 2; < gender > = 2;	

91) the conjugation-type translation pattern

reste toujours vivant dans	< number >	< gender >
pozostaje wiecznie żywy w	1	1
pozostaje wiecznie żywa w	1	2
pozostaje wiecznie żywe w	1	3
pozostają wiecznie żywi w	2	1
pozostają wiecznie żywe w	2	2
pozostają wiecznie żywe w	2	3
< case > = 6;		

92) the declination-type translation pattern

la mémoir des Français	< case >
pamięć Francuzów	1
pamięci Francuzów	2
pamięci Francuzów	3
pamięć Francuzów	4
pamięcią Francuzów	5
pamięci Francuzów	6
< number > = 1; < gender > = 2;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

93) the non-flexion-type translation pattern

en effet
istotnie

94) the declination-type translation pattern

les Français	< case >
Francuzi	1
Francuzów	2
Francuzom	3
Francuzów	4
Francuzami	5
Francuzach	6
< number > = 2; < gender > = 1;	

95) the conjugation-type translation pattern

connaissent	< number >	< gender >
zna	1	1
zna	1	2
zna	1	3
znają	2	1
znają	2	2

znają	2	3
< case > = 4;		

96) the declination-type translation pattern

les noms des Polonais illustre comme	< case >
nazwiska tak znanych Polaków jak	1
nazwisk tak znanych Polaków jak	2
nazwiskom tak znanych Polaków jak	3
nazwiska tak znanych Polaków jak	4
nazwiskami tak znanych Polaków jak	5
nazwiskach tak znanych Polaków jak	6
< number > = 2; < gender > = 3; < case > = 1;	

97) the declination-type translation pattern

Thadée Kościuszko	< case >
Tadeusz Kościuszko	1
Tadeusza Kościuszki	2
Tadeuszowi Kościuszce	3
Tadeusza Kościuszko	4
Tadeuszem Kościuszką	5
Tadeuszu Kościuszce	6
< number > = 1; < gender > = 1;	

98) the declination-type translation pattern

Frédéric Chopin	< case >
Fryderyk Chopin	1
Fryderyka Chopina	2
Fryderykowi Chopinowi	3
Fryderyka Chopina	4
Fryderykiem Chopinem	5
Fryderyku Chopinie	6
< number > = 1; < gender > = 1;	

99) the non-flexion-type translation pattern

et
i

100) the declination-type translation pattern

Marie Curie-Sklodowska	< case >
Maria Curie-Skłodowska	1
Marii Curie-Skłodowskiej	2
Marii Curie-Skłodowskiej	3
Marię Curie-Skłodowską	4
Marią Curie-Skłodowska	5
Marii Curie-Skłodowskiej	6
< number > = 1; < gender > = 2;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

101) the non-flexion-type translation pattern

ce sont là
to są
< case > = 1;

102) the declination-type translation pattern

des Polonais	< case >
Polacy	1
Polaków	2
Polakom	3
Polaków	4
Polakami	5
Polakach	6
< number > = 2; < gender > = 1;	

103) the conjugation-type translation pattern

qui se sont engagés aux côtés de	< number >	< gender >
który się zaangażował po stronie	1	1
która się zaangażowała po stronie	1	2
które się zaangażowało po stronie	1	3
którzy się zaangażowali po stronie	2	1
które się zaangażowały po stronie	2	2
które się zaangażowały po stronie	2	3
< case > = 2;		

104) the declination-type translation pattern

la France populaire	< case >
Francja ludowa	1
Francji ludowej	2
Francji ludowej	3
Francję ludową	4
Francją ludową	5
Francji ludowej	6
< number > = 1; < gender > = 2;	

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

105) the non-flexion-type translation pattern

mais
ale

106) the non-flexion-type translation pattern

pour les Français
dla Francuzów

107) the non-flexion-type translation pattern

c'est Adam Mickiewicz qui
to właśnie Adam Mickiewicz
< number > = 1;
< gender > = 1;

108) the conjugation-type translation pattern

est	< number >	< gender >
jest	1	1
jest	1	2
jest	1	3
są	2	1
są	2	2
są	2	3
< case > = 5;		

109) the declination-type translation pattern

l'incarnation vivante	< case >
żywe wcielenie	1
żywego wcielenia	2
żywemu wcieleniu	3
żywe wcielenie	4
żywym wcieleniem	5
żywym wcieleniu	6
< number > = 1;	
< gender > = 2;	

110) the non-flexion-type translation pattern

de son pays
swojego kraju

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

111) the declination-type translation pattern

Victor Hugo	< case >
Wiktor Hugo	1
Wickora Hugo	2
Wiktorowi Hugo	3
Wiktora Hugo	4
Wiktorem Hugo	5
Wiktorze Hugo	6
< number > = 1;	
< gender > = 1;	

112) the conjugation-type translation pattern

a dit que parle de	< number >	< gender >
powiedział że mówić o	1	1
powiedziała że mówić o	1	2
powiedziało że mówić o	1	3
powiedzieli że mówić o	2	1
powiedziały że mówić o	2	2
powiedziały że mówić o	2	3
< case > = 6;		

113) the declination-type translation pattern

Mickiewicz	< case >
Mickiewicz	1
Mickiewicza	2
Mickiewiczowi	3
Mickiewicza	4
Mickiewiczem	5
Mickiewiczu	6
< number > = 1;	
< gender > = 1;	

114) the non-flexion-type translation pattern

c'est parle
to mówić o
< case > = 6;

115) the declination-type translation pattern

du beau	< case >
piękno	1
piękna	2
pięknu	3
piękno	4
pięknem	5
pięknie	6
< number > = 1;	
< gender > = 3;	

116) the declination-type translation pattern

du juste	< case >
sprawiedliwość	1
sprawiedliwości	2
sprawiedliwości	3
sprawiedliwość	4
sprawiedliwością	5
sprawiedliwości	6
< number > = 1;	
< gender > = 2;	

117) the non-flexion-type translation pattern

et
i

118) the declination-type translation pattern

du vrai	< case >
prawda	1
prawdy	2
prawdzie	3
prawdę	4
prawdą	5
prawdzie	6
< number > = 1; < gender > = 2;	

119) the non-flexion-type translation pattern

c'est parle
to mówić o
< case > = 6;

120) the declination-type translation pattern

du droit	< case >
prawo	1
prawa	2
prawu	3
prawo	4
prawem	5
prawie	6
< number > = 1; < gender > = 3;	

121) the conjugation-type translation pattern

dont il a été le soldat	< number >	< gender >
którego był żołnierzem	1	1
której był żołnierzem	1	2
którego był żołnierzem	1	3
których był żołnierzem	2	1
których był żołnierzem	2	2
których był żołnierzem	2	3

122) the declination-type translation pattern

du devoir	< case >
obowiązek	1
obowiązku	2
obowiązkowi	3
obowiązek	4
obowiązkiem	5
obowiązku	6
< number > = 1; < gender > = 1;	

123) the conjugation-type translation pattern

dont il a été le héros	< number >	< gender >

ktòrego byl bohaterem	1	1
której byl bohaterem	1	2
ktòrego byl bohaterem	1	3
których byl bohaterem	2	1
których byl bohaterem	2	2
których byl bohaterem	2	3

124) the declination-type translation pattern

de la liberté	< case >
wolność	1
wolności	2
wolności	3
wolność	4
wolnością	5
wolności	6
< number > = 1; < gender > = 2;	

125) the conjugation-type translation pattern

dont il a été l'apôtre	< number >	< gender >
którego byl apostołem	1	1
której byl apostołem	1	2
którego byl apostołem	1	3
których byl apostołem	2	1
których byl apostołem	2	2
których byl apostołem	2	3

126) the non-flexion-type translation pattern

et
i

127) the declination-type translation pattern

de la délivrance	< case >
wyzwolenie	1
wyzwolenia	2
wyzwoleniu	3
wyzwolenie	4
wyzwoleniem	5
wyzwoleniu	6
< number > = 1; < gender > = 3;	

128) the conjugation-type translation pattern

dont il a été le précurseur	< number >	< gender >
którego byl prekursorem	1	1
której byl prekursorem	1	2
którego byl prekursorem	1	3
których byl prekursorem	2	1
których byl prekursorem	2	2
których byl prekursorem	2	3

At the beginning of a new sentence the value 1 is assigned to the <case> attribute.

< case > = 1;

129) the declination-type translation pattern

Victor Hugo et Adam Mickiewicz	< case >
Wiktor Hugo i Adam Mickiewicz	1
Wiktora Hugo i Adama Mickiewicza	2
Wiktorowi Hugo i Adamowi Mickiewiczowi	3
Wiktora Hugo i Adama Mickiewicza	4
Wiktorem Hugo i Adamem Mickiewiczem	5
Wiktorze Hugo i Adamie Mickiewiczu	6
< number > = 2; < gender > = 1;	

130) the conjugation-type translation pattern

symboliset bien	< number >	< gender >
dobrze symbolizuje	1	1
dobrze symbolizuje	1	2
dobrze symbolizuje	1	3
dobrze symbolizują	2	1
dobrze symbolizują	2	2
dobrze symbolizują	2	3
< case > = 4;		

131) the declination-type translation pattern

l'amitié	< case >
przyjaźń	1
przyjaźni	2
przyjaźni	3
przyjaźń	4
przyjaźnią	5
przyjaźni	6
< number > = 1; < gender > = 2;	

132) the non-flexion-type translation pattern

entre nos deux peuples
pomiędzy naszymi narodami

The result from the French-to-Polish machine translation system is the following Polish translation of the French input text:

Chopin urodził się w roku 1810 w Żelazowej Woli blisko Warszawy. Jego ojciec uczył tam języka francuskiego w bogatej rodzinie. Następnie jego rodzina przeniosła się do Warszawy. On uczył się gry na fortepianie i uczęszczał na zajęcia w Konserwatorium. Skoro tylko rozpoczął studia jego nauczyciele zauważyli że był genialnie uzdolniony. Kiedy ukończył swoje studia w Konserwatorium Warszawskim on opuścił Polskę i on wyjechał do Paryża gdzie on pozostał do końca swojego życia. On dawał tam najpierw prywatne lekcje gry na fortepianie i przeżywał tam ciężkie dni. Ale od chwili gdy dał swoje pierwsze koncerty on stał się sławny. On zawarł znajomość z powieściopisarką

George Sand z którą on pozostał w przyjaźni przez długie lata. Oni podróżowali razem na Majorkę. Stan zdrowia Chopina pogorszył się po tym jak wrócił do Francji. Choroba płuc na którą cierpiał pogłębiała się z dnia na dzień. Pomimo jego cierpień on nadal komponował arcydzieła które uczyniły go nieśmiertelnym. On umarł w Paryżu w roku 1849 myśląc o swojej ojczyźnie Polsce.

Polacy od dawna uczestniczyli w życiu Francji. I pamiątki po niektórych z nich pozostają wiecznie żywe w pamięci Francuzów. Istotnie Francuzi znają nazwiska tak znanych Polaków jak Tadeusz Kościuszko Fryderyk Chopin Maria Curie-Skłodowska. To są Polacy którzy się zaangażowali po stronie Francji ludowej. Ale dla Francuzów to właśnie Adam Mickiewicz jest żywym wcieleniem swojego kraju. Victor Hugo powiedział że mówić o Mickiewiczu to mówić o pięknie sprawiedliwości i prawdzie to mówić o prawie którego był żołnierzem obowiązku którego był bohaterem wolności której był apostołem i wyzwoleniu którego był prekursorem. Victor Hugo i Adam Mickiewicz dobrze symbolizują przyjaźń pomiędzy naszymi narodami.

The obtained Polish translation is both accurate and of high quality. It also sounds very natural and resembles the effect of the work of a human translator.

Conclusions:

In this paper the proposal by this author of a pattern-based machine translation technique is presented. The implementation of this technique for the system which translates from French to Polish is also described.

The obtained Polish translation is of high quality even if the length of translation patterns is not great. The translation patterns length statistics is presented in Fig. 1.
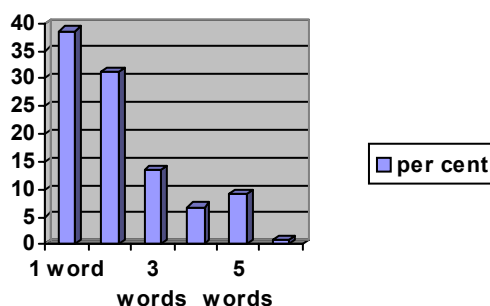

Fig. 1. The statistics of the translation patterns length.

As one can conclude from Fig. 1 most of the translation patterns have a length of one or two words. Exactly 38.6% translation examples have a length of one word, 31.1% translation patterns have a length of two words, 13.6% three words, 6.8% four words, 9.1% five words, and only 0.8% more than five words. This fact suggests convincingly that the pattern-based machine translation is aimed in the right direction. The final success of building a fully-automatic high-quality machine translation system, which would be able to translate unrestricted text from French to Polish depends on the amount of data gathered in the database, on which this author is currently working.

References:

[1] D. Arnold, L. Balkan, S. Meijer, R. L. Humphreys, L. Sadler: *Machine Translation: An Introductory Guide*, NCC Blackwell, London, 1994

[2] P. Whitelock, K. Kilby: Linguistic and Computational Techniques in Machine Translation System Design, UCL Press, London, GB, 1995

[3] C. Rico: *From Novelty to Ubiquity: Computers and Translation at the Close of Industrial Age*, http:\\www.accurapid.com/journal/15mt2.htm

[4] A. Melby: *Machine Translation and Philosophy of Language*; Machine Translation Review, No. 9, April 1999, pp. 6-17

[5] R. D. Brown: *Adding linguistic knowledge to a lexical example-based translation system*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA, 2000

[6] R. D. Brown: *Example-based machine translation in the Pangloss system*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA, 2000

[7] Y. Zhang, R. D. Brown, R. E. Frederking: *Adapting an example-based translation system to Chinese*, Carnegie Mellon University, Pittsburgh, PA, USA, 2000

Report on the 6th EAMT Workshop *Teaching Machine Translation*
November 14-15 2002: UMIST, Manchester, England

Organised by the
European Association for Machine Translation
in association with the
British Computer Society Natural Language Translation Specialist Group

by
Judith Belam and Derek Lewis
University of Exeter

Synopses of the Papers

Orie Miyazawa, University of Wolverhampton (*MT Training for Business People and Translators*), described how the PAT MT translation system (bi-directional English-Japanese) is used to provide in-house training for translators, managers and engineers at Epson Telford Ltd UK and to promote the use of machine translation. She outlined the types of documents used by the company and stressed the importance of pre- and post-editing in the translation process. Some basic rules for using a controlled language were also included.

Sivaji Bandyophadhyay, Jadavpur University (*Teaching MT: An Indian Perspective*), was unable to attend the workshop in person, but his paper was delivered by a member of UMIST. He drew attention to various factors affecting the teaching of MT, in particular the state of MT research in the country, language policy and patterns in the education system, student motivation and the status of the software industry in a multilingual setting. He described in detail the state of MT in India and referred to a number of institutions in which MT is included in the computer science or computational linguistics curriculum.

Elia Yuste, University of Zurich (*MT and the Swiss Language Service Providers*), outlined the linguistic situation in Switzerland with particular reference to the growing need for translation and multilingual documentation. Awareness and usage of computer mediated translation tools, however, proved to be surprisingly low in Switzerland. The author presented the results of a questionnaire distributed to language professionals and to service providers. They demonstrated the extent to which computer-based and CAT tools are employed in Switzerland and highlighted the need for awareness raising and training.

Dimitri Kalantzi, UMIST (*Teaching MT and Computer-Assisted Translation Tools in Greece*), described the results of her survey of teaching and training in MT/CAT in private schools and university departments in Greece. Although currently at relatively low levels, there are signs that education in MT is beginning to take off, mainly at undergraduate level. Obstacles to improvement include lack of government funding, low awareness levels, and lack of trained staff and established teaching posts; in addition, few Greek universities offer degrees in translation for undergraduates/postgraduates. The paper concluded with recommendations for change.

Paul Bennet, UMIST (*Teaching Contrastive Linguistics for MT*), presented a framework for teaching contrastive-oriented linguistics to students of machine translation. He included a classification of cross-linguistic differences and showed how such differences can be neutralised in abstract representations in order to simplify the operation of transfer in MT. He

described a canonical form for classifying language differences across languages and outlined how it can be used to identify diverging features such as movement, insertion and omission which can then be accommodated in formulating simplified transfer modules. Finally, he discussed the model's relevance for a broader typological foundation of language differences.

Andy Way, Dublin City University (*Testing Students' Understanding of Complex Transfer*), focussed on how to provide students with an understanding of the computational process of complex transfer in MT systems. Students are required to identify and classify instances of complex transfer, draw up dependency structures, write translation rules in a notation of their choice, and construct a Prolog-type rule which generates the translation from an input string. The methodology is appropriate for those teaching more advanced programming skills.

J. Gabriel Amores, University of Seville (*Teaching MT with Xepisteme*), described Xepisteme, a tool for developing transfer-based MT systems. Based on lexical functional grammar it is being used to teach MT at the universities of Seville and Pempeu Fabra in Spain. The tool's advantages are that it gives the user the control to monitor and manipulate each processing stage (analysis, transfer, or full translation) and to plug in different components (e.g. associate a lexicon with different grammars). With no programming expertise the user can also develop new modules, including lexicons and grammar processing rules. Plans are afoot to implement a MS Windows version, to provide a direct MT (as opposed to transfer) version, and to link the transfer module to WordNet.

Walther von Hahn and Christina Vertan, University of Hamburg (*Architectures of 'Toy' Systems for Teaching Machine Translation*), described how students, provided with a small, annotated lexicon (based on a 100-sentence test corpus), were given the task of designing and implementing an MT system and its various modules. Students also evaluated their systems and the educational success of the approach.

Svetlana Sheremeteyva, Handelshøjskolen i København (*An MT Learning Environment for Computational Linguistics Students*), argued for MT software specifically designed for training purposes. She presented the APTrans system (an experimental interactive MT system for translating patent claims between any pair of European languages) and its developer tools as a possible model.

Chi-Chinag Shei, National Tsing Hua University (*Teaching MT through Pre-editing*), outlined some of the major differences between Chinese and English, with special reference to areas of structural, lexical and semantic difficulty, and illustrated how students in Taiwan were encouraged to develop pre-editing skills in order to improve output from an English-Chinese MT system. The process not only familiarised students with MT technology but also promoted their competence in English composition.

Pointing to the growing global demand for post-editing, Sharon O'Brien from the Dublin City University (*Teaching Post-editing: A Proposal for Course Content*) discussed the skills which a post-editor requires and the need for post-editors to understand the technologies of MAT and also relationship between pre-editing and MT output. She outlined the elements of a proposed course in post-editing.

Enrique Torrejón and Celia Rico, Universidad Europea (*Controlled Translation: A New Teaching Scenario Tailor-Made for the Translation Industry*), discussed in some detail the concept of controlled languages, reviewed practical examples of CLs that have been used in industry, and outlined guidelines and writing rules. They described the methodology and types of practical exercise underpinning a new course at the European University in Madrid

which aims to teach the pre- and post-editing skills that will be increasingly required from professional translators.

The paper by Heather Fulford, University of Loughborough (*Freelance Translators and Machine Translation*), described the results of an exploratory study undertaken to establish the uptake of MT among freelance translators in the UK. Preliminary findings suggested that a relatively low uptake of MT was matched by a keen interest in learning more about MT and in seeing the development of resources that could provide self-directed training.

Mark Shuttleworth, Imperial College (*Combining MT and CAT on a Technology-Oriented Translation Masters*), reported on a project to evaluate how students on a masters-level MT programme performed in using a combination of MT and TM technologies to translate a large extract of a document on medical information. He concluded that the project gave students an important learning experience, although outcomes might depend on the particular software tools used (e.g. TRADOS v. Déjà Vu, and SYSTRAN) and on their interaction.

Judith Belam, University of Exeter (*Teaching MT Evaluation by Assessed Portfolio*), described her experiences of an assessed independent study project on MT evaluation for final-year undergraduates on a modern languages degree. After discussing the rationale of teaching MT, she presented several reasons why the project extended students in terms of motivation, competence in the foreign language, and transferable skills.

Mikel Forcada, University of Alicante (*Explaining Real MT to Translators: Compositional Semantics and Word-for-Word*), explored the advantages of presenting MT to students as a middle way between the theoretically motivated model of semantic compositionality and a set of refinements over basic word-for-word substitution. He aimed to influence students' expectations of MT by showing them the differences between the compromises required to implement a linguistically accurate system and the volume of work needed in order to bring a word-for-word model (model zero) to reasonable levels of performance.

Federico Gaspari, UMIST (*Using Free On-line Services in MT Teaching*), discussed the role that free on-line machine translation services can play in teaching MT. After outlining their advantages (freely available in many language pairs) and disadvantages (no access to rules or dictionaries for customisation), he focuses on the increasing need for web localisation and multilingual on-line content management. In conclusion he proposed a broad approach in teaching MT to students which emphasises the role in communication that on-line MT can play rather than its technical or linguistic performance.

Natalie Kübler, University of Paris 7 (*Teaching Commercial MT to Translators: Bridging the Gap between Human and Machine*), presented an experiment conducted at the University of Paris which aimed to show translation trainees how and why MT can be incorporated into their work. She described the translation projects which the students undertook and the language data and tools which they used in carrying the projects through to submission to the customer. The data comprised specialised dictionaries and corpora (generally available on a web-based interface) and the tools included terminology extraction software, a concordancer, and MT (Systran). Students worked in groups to create customised dictionaries, evaluate the MT system's performance and post-edit the results. As a result students were better able to grasp the advantages of MT and appreciate why general purpose MT is hard to achieve. They were also able to place MT in context alongside human translation.

Who Uses MT?

Apart from its orientation towards teaching MT the workshop also offered some answers to the question of who actually uses MT in practice. After a couple of sessions with a proprietary MT system it might be tempting to conclude that no-one could ever seriously use MT. However, the workshop demonstrated conclusively that such a conclusion would be premature. Orie Miyazawa described how a UK subsidiary of the Japan-based Epson Company, Epson Telford, manufactures ink cartridges for consumer inkjet printers, with all manufacturing methods and procedures originating from the headquarters in Japan. Senior positions in the company are occupied by Japanese staff from the head office, some of whom require assistance in communication. The translation requirement is therefore extremely high and the requirements cannot always be met by human translation. The author stressed the importance of training translators in the use of MT, the importance of pre-and post-editing, and devoted a long section to discussing the type of documents and their suitability for translation by MT. She identified some document types as being unsuitable (e.g. presentations made for senior board members from Japan, minutes of meetings) and other as being eminently suitable (e.g. parts lists; with often a simple word-for-word translation requiring entry of the names of the parts into the dictionary). In other cases (e.g. engineering instructions) the decision was unclear: on the one hand, accuracy is vital, while on the other hand the material often need to be translated very quickly. With regard to e-mails, some need thorough translation if the information is very important; while others include communications about company events and information, which is generally simple and straightforward and does not require a high quality translation, especially since all the addressees are members of the company.

In conversation with Viggo Hansen, whose company, Zacco A/S (Hellerup, Denmark), translates patents into Danish, it emerged that a patent becomes valid in Denmark only after it is translated into Danish. Naturally this generates a large and on-going requirement for large amounts of translation. Mr Hansen, who was required to translate approximately 3 million words a year, explained that he valued the speed and consistency of MT for this task, and stressed the ease of post-editing predictable MT errors compared with the difficulty of correcting human translation with its unpredictable errors, misinterpretations and omissions.

In the light of the above it might be expected that MT is becoming more widespread. There were, however, some surprises here too. Elia Yuste (University of Zurich) had begun a survey of the translation industry in Switzerland, where one might expect the large translation requirement in a multilingual nation to generate a large demand for MT. Surprisingly, perhaps, her survey has shown so far that, 'with the exception of two leading corporate language service providers who have performed evaluation exercises of MT systems and adopted one in their workflow, there is no overall interest in MT in the Swiss translation arena. Regrettably, fears and misconception about MT still remain'.

In India, too, it emerged from Sivaji Bandyopadhyay's paper that the use of MT does not seem to be widespread, although he stressed the need for it: According to the author India is a vast multilingual country where there are 18 scheduled languages. There is a great demand for translation of documents from one language to another. Most of the state governments work in the respective regional languages whereas the Union Governments official documents and reports are in bilingual form (Hindi/English). In order to have a proper communication there is a need to translate these reports and documents into the respective regional languages. With the limitations of human translators most of this information is missing and is not percolating down. A machine assisted translation system or a translators' workbench would increase the efficiency of human translators.

MT and Professional Translators

Another interesting question was the degree to which professional translators are informed about MT. The survey by Heather Fulford (Loughborough University Business School) suggested that only 7% were actively using MT, although over half had been asked to do post-editing work. There was a general feeling that more training in MT would be useful to them.

Many of the speakers at the workshop were engaged in training translators, and there were proposals about what trainee translators should be learning. Several areas came up for consideration. One of these was *pre-editing and controlled languages.* Enrique Torrejòn and Celia Rico from the Universidad Europea de Madrid talked about the importance of controlled languages and the fact that their use is becoming more widespread, thus giving rise to a whole new area of professional translation which the authors called 'controlled translation'. In their words: 'Pre-editing texts following the constraints from the controlled language specifications or even writing texts directly in CL become a new set of skills that translators need to take into account.' Chi-chiang Shei stressed the importance of pre-editing when translation is required between two languages whose structure differs very markedly, such as English and Chinese.

A second area was that of *post-editing*. Sharon O'Brien from Dublin City University proposed including a component in training post-editors into a course for translators. She underlined the value of post-editing as a way of making the most of the speed and efficiency of MT, quoting a study by Vasconcellos and Léon which claims that 'a full-time trained post-editor working on-screen, can produce polished, standard quality output at a rate of between two and three times faster than traditional translation'. She went on to state that post-editing was a skill which needed to be learned and gave some suggestions for course content.

A third area had to do with *understanding how an MT systems works* at a technical level. Andy Way from Dublin City University described the need for all students of MT to have an idea of how the system actually operates. His own students were enrolled on a three-year course in computational linguistics, although he pointed out that 'Any contemporary course on MT ought to equip students with knowledge of the differences between rule-based and statistical MT; direct and indirect approaches; and transfer-based and interlingual systems.' He went on to talk about the issue of complex transfer, 'an integral component of this section of a course on MT' and which referred to those difficult cases of translation where the grammatical structure diverges considerably between languages (as in German *Es gefällt mir* for English *I like it*).

Chan Sin-wai (ed.) (2002) *Translation and Information Technology*, The Chinese University of Hong Kong: Chinese University Press, ISBN 962-996-077-X, Price US$18

Introducing this book, the editor, who heads the Department of Translation at the Chinese University of Hong Kong, draws attention to the fact that MT systems designed by computational linguists and focusing on formalistic analyses of language have so far produced rather disappointing results. He argues for a greater reference to 'language in use' rather than a continuing emphasis on grammatical models and suggests that practising translators will in future become more closely involved in MT development. Whether the collection of thirteen papers that follows bears out these points is for the reader to decide, but it covers a fascinating range of systems and work in progress in the field of MT. Particular attention is paid to systems involving Chinese and under development in China.

The volume is divided into three thematic sections: methodology and application, terminology, and critique and training.

The first section describes two Chinese-English MT systems. In the first of these Chan Sin-wai presents TransRecipe, an MT program which combines corpus-based, example-based, pattern-based and rule-based approaches in a single hybrid system designed to translate Chinese cooking recipes into English. The second system (presented by Lin Quen of Peking University) likewise uses a 'multi-engine' architecture for general-purpose translations. In both cases it is the blending of approaches which reflects a re-orientation away from formal-theoretical paradigms towards strategies which more closely model the work of the practising translator. While Chang Baobao et al describe work on constructing and designing a Chinese-English bilingual corpus for use in translation memory systems, a paper by Li Tangqiu et al present a hybrid method for disambiguating syntactic structures and word senses in Chinese which chooses the best candidate the best candidate from multi-parse trees by comparing the degree of syntactic and semantic collocation between the words in the structure being analysed. The knowledge to do this comes from word entries stored in Hownet, a knowledge base specifically designed for NLP. In a contribution on example-based machine translation (EBMT) Kit Chunyu et al provide a comprehensive, yet concise and informative overview of the historical development of example-based MT (EMBT) and the methods used. The paper would be an excellent starting point for students needing to identify the basic issues and able to research further the algorithmic techniques that are briefly presented. Jonathan Webster et al follow this up by presenting a project at the City University of Hong Kong which is applying the example-based approach to meet the growing demand for English-Chinese translations of legal documents in Hong Kong. For data acquisition and alignment purposes semantic web technology is used to recode a 25 million word bilingual (English-Chinese) legal document corpus into marked-up, machine-readable form (in RDF/XML format).

In a section on terminology Aman Chiu and Björn Jernudd review the practical difficulties which translators and other specialists have to contend with in identifying and standardising terminological equivalents in English and Chinese with reference to the language of IT. A major problem is the existence of at least four standardisation bodies which appear not to operate with guidelines or criteria for the identification, collection and maintenance of terminological data. Pursuing this theme further, Charlotte To and Björn Jernudd discuss the terminological problems encountered by professional users of the internet (technical jargon, acronyms, pronunciation, spelling, meaning and usage) in a range of situations (business meetings, conversations, seminars, reading technical manuals, writing articles for publication

on the web). They outline the strategies which users bring to bear in order to overcome obstacles of comprehension and translation.

Turning to training issues, Beverly Adab provides a detailed and informative insight into how a wide variety of software resources (notably corpora, of which the internet provides the largest single source, and search tools/concordancers such as Wordsmith and Multiconcord) are used in a postgraduate translation programme at the University of Aston. Carrie Chau Kam Hung and Irene Ip Kwok Chun report on a survey in which they assessed the perceived benefits to students of translation and interpreting of using a computer-based vocabulary-learning programme. A paper by Li Defeng provides rather negative feedback on the comprehensibility of texts generated by some well-known English-Chinese MT programs, although he also emphasises the need to develop better software and the importance of incorporating a range of IT resources in training. Paris Lau Chi-chuen supplies some interesting insights into the Chinese tradition of translation and asks whether new technology has resulted in the values of the traditional human translator being supplanted by the need for rapid decoding, speed of comprehension, partial fidelity and overall intelligibility. Comparing the translation by three Chinese-English MT systems of a very short comment by an anonymous MT researcher on the relationship between language and literature, science and machine translation, Evangeline S. P. Almberg provides an ironical comment on what computers must yet achieve before they can match the expert human translator.

In summary this volume, which unlike so many on MT is reasonably priced, provides informative perspectives on research, development, terminological and training issues. We can also expect some major advances in MT resources from China in the near future.

Derek Lewis

University of Exeter

Conferences and Workshops

The following is a list of recent (i.e. since the last edition of the MTR) and forthcoming conferences and workshops. E-mail addresses and websites are given where known.

A comprehensive calendar of conferences is maintained at the EACL website:
http://www.coli.uni-sb.de/eacl/calendar.php3

28 March – 1 April 2002: Corpus Linguistics 2003
Lancaster University, UK
http://www.comp.lancs.ac.uk/ucrel/cl2003/
E-mail: cl2003@comp.lancs.ac.uk or mcenery@comp.lancs.ac.uk

16–22 February 2002: CICLing2003
Fourth International Conference on Intelligent Text Processing and Computational Linguistics
Mexico City, Mexico
http://www.cicling.org/2003

23–25 April 2002: IWPT 2003
8th International Workshop on Parsing Technologies
Nancy, France
http://iwpt03.loria.fr/

ESSLLI-2003: 18– 29 August 2002
15th European Summer School in Logic, Language and Information
Vienna, Austria
http://www.esslli2002.it

24 August – 1 September 2002: COLING 2002
Taiwan
http://www.coling2002.sinica.edu.tw/

9–12 September 2002: TSD 2002
Fifth International Conference on TEXT, SPEECH and DIALOGUE
Brno, Czech Republic
http://www.fi.muni.cz/tsd2002/

8–12 October 2002: AMTA 2002
The Association for Machine Translation in the Americas
Tiburon (near San Francisco), California, USA
http://www.amtaweb.org/AMTA2002/

4–7 November 2002: LISA Forum Europe: Standards in Localization and Translation
Heidelberg, Germany
E-mail: lisa@lisa.org
http://www.org/events/2002europe

12 November 2002: IBERAMIA Workshop: Multilingual Information Access and Natural Language Processing
Seville, Spain
E-mail: anselmo@lsi.uned.es
http://www.sensei.lsi.uned.es/iberamia-mlia

14–15 November 2002: 6th EAMT Workshop: Teaching Machine Translation
UMIST, Manchester, UK
E-mail: Harold.Somers@umist.ac.uk
http://www.ccl.umist.ac.uk/events/eamt-bcs

21–22 November 2002
ASLIB: 24th Translating and the Computer Conference
London, UK
E-mail: barbara.hobbs@aslib.com
http://www.aslib.com/conferences

29 November 2002: CLIN 2002: Computational Linguistics in the Netherlands
E-mail: clin@let.rug.nl
http://www.odut.let.rug.nl/clin2002

22–25 March 2003: AAAL (American Association for
Applied Linguistics)
Sheraton National Hotel in Arlington, VA, USA
http://www.aaal.org/aaal2003/

11 April 2003: COMPLEX 2003
7th Conference on Computational Lexicography and Text Research
Budapest, Hungary
http://www.conferences.hu/EACL03/complex.htm

7–10 May 2003: World Call 2003
Banff, Alberta, Canada
http://www.worldcall.org/

12–17 April 2003: EACL 2003
10th Conference of the European Chapter of the Association for Computational Linguistics
Budapest, Hungary
http://www.conferences.hu/EACL03/
For the workshop on 'MT and other language technology tools; improving MT through other language technology tools; resources and tools for building MT' contact Bente Maegaard, e-mail: bente@cst.dk, http://www.elsnet.org/workshops/w01.doc

15–17 May 2003
7th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop: Controlled Language Translation
Dublin City University, Ireland
E-mail: Andy Way (away@computing.dcu.ie)
http://www.eamt.org/eamt-claw03/

7–12 July 2003: ACL2002
41st Annual Meeting of the Association for Computational Linguistics
Sapporo Convention Center, Sapporo, Japan
http://www.ec-inc.co.jp/ACL2003/

7–9 July 2003: ITHET 2003
4th International Conference on Information Technology Based Higher Education and
Training
Marrakech, Morocco
http://www.emi.ac.ma/ithet03

10-11 September 2003: Recent Advances in Natural Language Processing (RANLP2003)
Borovetz, Bulgaria
http://www.lml.bas.bg/ranlp2003

23-28 September 2003: Machine Translation Summit IX
New Orleans, USA
http://www.mt-summit.org

MEMBERSHIP: CHANGE OF ADDRESS

If you change your address, please advise us on this form, or a copy, and send it to the following
(this form can also be used to join the Group):

Mr. J.D.Wigg
BCS-NLTSG
72 Brattle Wood
Sevenoaks, Kent TN13 1QU
U.K.   Date: ....../....../......

Name: ....
Address: ....
....
Postal Code: .... Country: ....
E-mail: .... Tel.No: ....
Fax.No: ....

Note for non-members of the BCS: your name and address will be recorded on the central computer records of
the British Computer Society.

Questionnaire

We would like to know more about you and your interests and would be pleased if you would complete as much
of the following questionnaire as you wish (please delete any unwanted words).

1.
   a. I am mainly interested in the computing/linguistic/user/all aspects of MT.
   b. What is/was your professional subject? ....
   c. What is your native language? ....
   d. What other languages are you interested in? ....
   e. Which computer languages (if any) have you used? ....

2. What information in this Review or any previous Review have you found:
   a. interesting? Date ....
       ....
       ....
   b. useful (i.e. some action was taken on it)? Date ....
       ....
       ....

3. Is there anything else you would like to hear about or think we should publish in the *MT Review*?
       ....
       ....
       ....
       ....

4. Would you be interested in contributing to the Group by,

   a. Reviewing MT books and/or MT/multilingual software
   b. Researching/listing/reviewing public domain MT and MNLP software ....
   c. Designing/writing/reviewing MT/MNLP application software ....
   d. Designing/writing/reviewing general purpose (non-application specific) MNLP ....
       procedures/functions for use in MT and MNLP programming ....
   e. Any other suggestions? ....
       ....
       ....
       ....
                          Thank you for your time and assistance.