MACHINE TRANSLATION REVIEW

The Periodical
of the
Natural Language Translation Specialist Group
of the
British Computer Society
Issue No. 5
April 1997

The *Machine Translation Review* incorporates the Newsletter of the Natural Language Translation Specialist Group of the British Computer Society and appears twice yearly.

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
Department of German
Queen's Building
University of Exeter
Exeter
EX4 4QH
United Kingdom

Tel.: +44 (0)1392 264330
Fax: +44 (0)1392 264377
E-mail: D.R.Lewis@exeter.ac.uk

ISSN 1358–8346

Contents

Group News and Information 4
Letter from the Chairman 4
The Committee 5
BCS Library 5

The AMALGAM Parts-of-Speech Tagger 6
Roger Harris

Language Engineering Systems by Lingvistica ’93 and ETS Publishers Ltd:
English, Russian, Ukrainian 10
Michael Blekhman

The Ergo Parser Challenge 19
J.L. Morris

The Telegraph and Systran Machine Translation Systems for Personal Computer:
NLTSG Seminar 25
Derek Lewis

Conferences and Workshops 27

Membership 30

Group News and Information

Letter from the Chairman

The Group is now entering its third year of publication of the *Review,* a very creditable performance due to Derek Lewis's efforts as Editor and the voluntary contributions of papers and information supplied by members. However, as I have intimated previously, although we are generously supported by the British Computer Society, publishing twice yearly is now eating into our reserves and we will have to consider our future options soon.

One route is to upgrade the *Machine Translation Review* into a refereed Journal which would justify charging and collecting from its subscribers a worthwhile sum. Apart from recruiting qualified referees, this would create more work and require the present Committee to be enlarged with willing volunteers to manage the paperwork. We would, of course, need some assurance that sufficient quality papers would also be forthcoming. If any members have any views on this proposal please let me or any other Committee member know.

Another option might be to join with another organisation with similar or related interests with whom we could publish jointly. In the short run it has been suggested that we publish on our web pages at the BCS, which Roger Harris maintains so well for us, with the option of members being able to ask for printed copies at cost.

In any event we would welcome more articles, papers and reports on the subject of machine translation and related subjects such as computer assisted language teaching, computer based dictionaries and aspects of multilinguality in computing etc. We would welcome papers from staff and students in linguistics and related disciplines, and from translators and any other users of MT software.

I am also interested in reviews of some of the translation software being published on the Internet and I thank John Morris for his paper on the 'Ergo Parser Challenge' and Roger Harris for his paper on the collection of grammatical text taggers at Leeds University. I am still looking for someone to review the Link parser offered by Carnegie Mellon at http://bobo.link.cs.cmu.edu/grammar/html/intro.html, which looks interesting.

If you are sufficiently interested in machine (assisted) translation to read the *Review*, you could well have some interesting knowledge or experiences to pass on to other members, so please do not be backward in coming forward with further contributions.

Perhaps I could remind members that they do not need to live near London to assist the Committee. Although we do not have sufficient funds to pay travel expenses for all Committee members to attend meetings, we still welcome Correspondent members. Anyone interested in helping should contact me or any other Committee member.

I would also like to remind you that there is a lot of MT related information on our web pages at the BCS at http://www.bcs.org.uk/siggroup/sg37.htm.

I regret I have to apologise again for the non-appearance of the Proceedings of the International Machine Translation Conference at Cranfield in 1994, but Douglas Clarke now has usable copies of all the papers and is assembling the final product.

All opinions expressed in the *Review* are those of the respective writers and are not necessarily shared by the BCS or the Group.

J.D.Wigg

The Committee

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

David Wigg (Chair)	Tel.: +44 (0)1732 455446 (H)
	Tel.: +44 (0)171 815 7472 (W)
	E-mail: wiggjd@vax.sbu.ac.uk
Monique L'Huillier (Secretary)	Tel.: +44 (0)1276 20488 (H)
	Tel.: +44 (0)1784 443243 (W)
	E-mail: m.lhuillier@vms.rhbnc.ac.uk
Ian Thomas (Treasurer)	Tel.: +44 (0)181 464 3955 (H)
	Tel.: +44 (0)171 382 6683 (W)
Derek Lewis (Editor)	Tel.: +44 (0)1404 814186 (H)
	Tel.: +44 (0)1392 264330 (W)
	Fax: +44 (0) 1392 264377
	E-mail: d.r.lewis@exeter.ac.uk
Catharine Scott (Assistant Editor)	Tel.: +44 (0)181 889 5155 (H)
	Tel.: +44 (0)171 607 2789 X 4008 (W)
	E-mail: c.scott@unl.ac.uk
Roger Harris (Rapporteur)	Tel.: +44 (0)181 800 2903 (H)
	E-mail: rwsh@dircon.co.uk
Correspondent Members:	
Gareth Evans (Minority Languages)	Tel.: +44 (0)1792 481144
	E-mail: g.evans@sihe.ac.uk
Ruslan Mitkov	Tel: +44 (0)1902 322471 (W)
	Fax: +44 (0)1902 322739
	E-mail: R.Mitkov@wlv.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)171 240 1871; fax: +44 (0)171 497 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 to 5.00 pm.

Website

The website address of the BCS-NLTSG is: http://www.bcs.org.uk/siggroup/sg37.htm

The AMALGAM Parts-of-Speech Tagger

by

Roger Harris

Applying parts-of-speech tags to text is easy if you do it the AMALGAM way. If you can can send and receive an e-mail message then all you need to do is set a few simple parameters and then send your text to the automated POS tagger at Leeds University. Your text will be returned to your electronic mailbox fully tagged and within a few minutes .

The tagger is an experimental version and has been designed to operate on text written in English. It is based upon the Brill tagger and a detailed description is available at the 'General information' address below. AMALGAM is an acronym standing for **A**utomatic **M**apping **A**mong **L**exico-**G**rammatical **A**nnotation **M**odels.

The output varies according to which tag-set one has chosen and for any tag-set AMALGAM is likely to produce some wrong results. AMALGAM is designed to parse English sentences and it appears to be correct most of the time. AMALGAM will also tag a text in any other language in the world, although the results will all be wrong. As with so many things computational it requires a human expert to determine whether the output results are correct, co-incidentally correct or wrong.

No doubt you will want to test AMALGAM. Here are the various Internet addresses which you will need.

General information (more than 70 Kbytes):
URL: http://agora.leeds.ac.uk/amalgam/

Descriptions of the eight POS tag-sets (more than 240 Kbytes):
URL: http://agora.leeds.ac.uk/amalgam/tagsets/tagmenu.html
The POS tag codes are listed in the files.

E-mail address of the automatic AMALGAM POS Tagger:
E-mail: amalgam-tagger@scs.leeds.ac.uk

E-mail address for a friendly human response:
E-mail: sean@scs.leeds.ac.uk

To try out the system, send an ASCII-format e-mail message to the following address:
amalgam-tagger@scs.leeds.ac.uk
Subject: token verbose brown ice llc lob parts pow sec upenn
Message: When will the train depart for Edinburgh?

Within a few minutes you should receive eight messages, one for each POS tag-set. The line in the message field will be displayed with the words one below the other and with a POS tag alongside each word.

A help file may be obtained by sending a message to the Tagger address listed above. Put 'help' in the subject line and leave the message blank.

Before sending a text, the parameters for tagging must be set, not in the text itself but in the e-mail subject line. The parameter keywords are not case-sensitive. The text must be in ASCII format.

The parameters are the following:

A. Eight POS tagging systems:

Name:	**Tag-set parameter**
1) Brown Corpus	Brown
2) International Corpus of English	ICE
3) London-Lund Corpus	LLC
4) Lancaster-Oslo/Bergen Corpus	LOB
5) UNIX parts	Parts
6) Polytechnic of Wales Corpus	POW
7) Spoken English Corpus	SEC
8) University of Pennsylvania Corpus	UPenn

B. Concise or verbose reporting mode:
Parameter: verbose, noverbose.

C. Tokenise on or off:
Parameter: token, notoken.

The AMALGAM POS Tagger tends to send an output file between twice and twelve times the size of the input file. A large part of the output files appear to contain blank spaces (ASCII 32). It is suggested that the input file should not exceed 50 Kbytes. The system was designed to operate on English texts, so to test it I sent a small file containing a few short English sentences.

As a computer programmer I was curious to see what the effect of illegal input data might be, so, ignoring the English-only stipulation, the input file also contained sentences from 25 different foreign languages chosen at random from *Languages of Asia and The Pacific* by Professor Charles Hamblin (ISBN 0-207-15880-0). I included a sentence in Inuit (inuk kaagami iqalut siuriaqpuq) taken from 'Eskimo Stories' by Nungak and Arima, some sentences containing nonsense words, mispellings and repetitions, and a sentence (w aknwlj rzt v y ltr) from *Agili Writing* by Anne Gresham (ISBN 1-872968-00-7). The input file size was about 1.5 Kbytes and the output files were about twelve times larger.

The AMALGAM POS Tagger attached POS grammatical tags to all the words, English and non-English alike and only a few of the latter were actually tagged as foreign. That the non-English words were tagged was disconcerting to say the least. I expected the tagger to flag them as 'unknown' or something like that.

Typical output for a non-English sentence (illegal input):
Language: New Guinea Pidgin
Sentence: Yu laik wanem samting? (What do you want?)

Word	Brown	UNIX Parts	ICE
Yu	/NN noun, singular, common	/adj	/N(com,sing)
laik	/NN noun, singular, common	/adj	/N(com,sing)
wanem	/NN noun, singular, common	/noun	/N(com,sing)
samting	/VBG verb, present partic	/gerund /noun	/V(montr,ingp)
?	/. sentence terminator	.	?/PUNC(qm)

Non-English words formed about two thirds of the sample text. Very few of these (e.g. 'le,' 'mea' and '(sic)') were tagged as foreign. Perhaps the Brown Tagger recognised 'le' as French; it is actually also a Samoan word.

It may be argued that it was unfair of me to expect the tagger to handle sentences for which it was not designed. But if it allocates grammatical tags to all words regardless of their language, then how can one be sure that it is reliable for English alone? The tagger makes a calculated guess when it does not know something but does not tell one when this happens.

Typical output for an English sentence (legal input):
Sentence: The lamb chased after the wolf and leapt upon its back.

Word	Brown	UNIX Parts	ICE
the	/AT article	/art	/ART(def)
lamb	/NN noun, singular, common	/noun	/N(com,sing)
chased	/VBN verb, past participle	/verb	/V(intr,past)
after	/IN preposition	/prep	/PREP(ge)
the	/AT article	/art	/ART(def)
wolf	/NN noun, singular, common	/noun	/N(com,sing)
and	/CC conjunction, coordinating	/conj	/CONJUNC(coord)
leapt	/VBD verb, past tense	/noun	/N(com,sing)
upon	/IN preposition	/prep	/PREP(phras)
its	/PP$ determiner, possessive	/pos	/PRON(poss,sing)
back	/RB adverb	/adj	/ADV(phras)
./.	sentence terminator	./.	./PUNC(per)

When dealing with English words the tagger was somewhat better, although 'anent,' an English preposition, was tagged as a noun while 'light' ('... light the candle ...') was tagged as a noun and as an adjective. Two printing terms, 'shrdlu' and 'etaoin', were both tagged as nouns and adjectives. See also 'leapt' and 'back' in the previous table.

The need to reject or flag spurious, out-of-range or unrecognised input data is crucial in designing computer software but it may not be altogether straightforward to establish the language in which a text is written from the text itself. Even if the input language is specified as English the text may contain non-English words which may confuse the software. Imagine a text in which a gourmet discusses in English the foods and wines of many countries and names each in the local vernacular. Most of the words may be in English but there will be a high proportion of non-English words from many languages.

It would be interesting to know whether the tagging of an illegal word sets off a ripple of incorrect POS tags amongst nearby legal words whose POS tags are based upon the POS tag of the illegal word. If it did, how could one tell?

I feel uneasy in making these critical comments. The AMALGAM team, led by Eric Atwell, comprises only a handful of people and yet it has devised a large complex on-line system which operates smoothly and, for its intended input language, fairly reliably. The system is still at the prototype stage. As a programmer I know how difficult it can be to encode subtle and elusive concepts. Really, I should compliment the AMALGAM team on their efforts and I do.

Making the AMALGAM Tagger available in an easy-to-use on-line format should stimulate interest in taggers and in software which will use their output.

The tagger will parse anything from short sentences to million-word texts although it is preferred that large files are, by arrangement, submitted out of office hours to avoid congestion. Once the output file is returned it is likely to require further processing. A program might be purchased or written in order to extract the words and their associated tags from the output file. Writing such software, including dealing with eight different tag-sets, would be an interesting database problem. One would have to do a lot of careful work to get the extraction program to function correctly. One would still be faced with the problem of what to do with the data.

As an aid for language teachers and students, especially children, it should be marvellous. First, parse a sentence, then send the sentence to the tagger. One's work may be checked and lessons learned in a few minutes, assuming that the output is correct.

Roger Harris may be contacted at rwsh@dircon.co.uk

Language Engineering Systems by Lingvistica ’93 and ETS Publishers Ltd: English, Russian, Ukrainian

by

Michael S. Blekhman,
Director, Lingvistica ’93 Company
Head of the Laboratory for Machine Translation, Kharkov State Polytechnical University

Introduction

After 1991, when Ukraine declared its national sovereignity in the wake of the disintegration of the former Soviet Union, journalists in this part of the world, especially the politically engaged ones, insisted that the Ukraine and Russia could not exist on their own: vital economic links would be ‘torn apart’, people would be ‘isolated’ from each other, and ‘scientific life’ would ‘come to an end’.

I hope it will be interesting for people in the West to know the real situation — at least, in the field of language engineering. Being a professional linguist, or, to put it more precisely, language engineer, I would like to describe the fruits of the collaboration between two companies, the Ukrainian Lingvistica ’93 and the Russian concern ETS Publishers Ltd. This paper is intended to provide information and facts, not to persuade or to engage in advertising. Hopefully, it will serve as an introduction and stimulus to future discussion. The idea of writing the paper came to me after reading a brilliant paper by Derek Lewis in *Machine Translation Review* (No.4, 1996), which describes the Power Translator German-English system. I hope my paper will be as informative and interesting to read as that by D. Lewis.

I would be happy to hear opinions and to respond to questions. Later this year, I also hope to provide detailed description of the German-Russian MT system being developed by Lingvistica ’93 and ETS Ltd.

Background

Lingvistica ’93 Co. and ETS Publishers Ltd. have established a fairly solid position in the market of language engineering products of the former Soviet Union. These companies are headed by M. S. Blekhman and I. V. Fagradiants, respectively.

I have the honour of being a pupil of one of the most outstanding Russian linguists, Professor Raimund Piotrowski, who has trained and nurtured dozens of language specialists (including, among others, the authors of the well-known Stylus translation system). During the twenty years of my own professional activity we have developed computer-based systems for information retrieval, abstracting, indexing, and, of course, machine translation. I am also pleased to mention Professor Victor Berzon and Dr. Boris Pevzner as my teachers: the former was one of the most authoritative specialists in discourse analysis in the former Soviet Union; the latter was the first to formulate the idea of example-based machine translation in this country (in the early 1970s!).

My friend and partner Igor Fagradiants founded a unique publishing house in Moscow, ETS Publishers Ltd, for producing electronic and traditional dictionaries in book form. In

conjunction with other specialists, Igor has developed a series of Finnish-Russian (!) dictionaries, whose high quality is greatly appreciated by his Finnish colleagues.

This paper gives a detailed description of some of our products. The paper calls a spade a spade: the last thing I want is to persuade the reader that these systems are perfect, or even that we have solved the principal language engineering problems. Nothing of the kind. I will try to give readers objective information — and let them draw their own conclusions.

PARS-PARS/U-RUMP: a Three-Language MT Package

Functional Description

In the Ukraine three languages are widely used, though in different capacities. Russian is the native tongue of the overwhelming majority of people living in towns and cities. Ukrainian is the official language and its usage is increasing; its status and prospects are to some extent similar to that of Hebrew in Israel. English is the language of international communication: it is employed, for instance, on the Internet, on distributed CDs, and in technical documentation.

Since 1986 we have been developing the English-Russian-English PARS system (the title is the abbreviation or acronym of the Russian name denoting 'Translating English and Russian Papers'). In addition, since 1990, we have been developing RUMP (meaing: 'Russian-Ukrainian-Russian Machine Translation'). PARS is currently marketed in Russia by ETS Publishers Ltd. The CD-ROMs comprise PARS and/or the Polyglossum system of dictionaries. These products have become the most popular translation systems in the huge Russian market: more than 10,000 CDs were sold between the end of 1995 and April 1997. In 1996 another system by Lingvistica '93 appeared on the market: called PARS/U, this translates between English and Ukrainian. All three systems are quite similar; so having mastered, for example, RUMP, the user would easily learn how to use PARS and PARS/U. Each system runs in either Windows or DOS.

PARS, RUMP, and PARS/U: the DOS Versions

One of the main peculiarities of these systems is the user(translator)-oriented **built-in two-window editor**. It features some specific functions that correspond to the most frequent text-editing operations performed by professional translators:

- key-stroke transposition of neighbouring words
- key-stroke change of register (substitution of uppercase letters with lower case, and vice versa)
- marking polysemantic words and phrases in the target text with asterisks; the user may easily substitute a translation variant
- search for the next 'new' word, i.e. a word not found in the dictionary
- the possibility of entering 'new' words into the dictionary directly from the text editor, according to the principle 'dictionary first': the user opens the dictionary and initiates the procedure for entering the next 'new' word while the word entered is highlighted in the text; this means that the user can see the context of the word and insert the right translation.

The screen may be split either horizontally or vertically, and the user may scroll either both windows synchronously, or the active one only. In addition the target text may be exported to

another text editor supporting ASCII files, such as PenEdit, a pen editor developed by the Kiev-based team led by Dr. Alexander I. Kazakov.

PARS, RUMP, and PARS/U: Windows versions

These systems work under Windows 3.1 and Windows 95. They translate files in formats such as WinWord, HTML, as well as Windows Help-files.

Each system may be activated directly from the wordprocessing application MS Word 6.0 or MS Word 7.0: once the MT systems have been installed, the main menu of the application will include the item 'Translate', with the option of selecting the corresponding system PARS, PARS/U, or RUMP. The user opens the source text in the editor and starts up one of the MT systems, after which the machine translation of the text appears in the lower window created by MS Word; the translated target text preserves the formatting of the source text, i.e. features such as fonts, styles, and tables. The polysemantic words and phrases are marked with asterisks, as in the DOS versions.

Figure 1 shows how PARS/U has translated the Declaration of State Sovereignity of the Ukraine from Ukrainian into English.

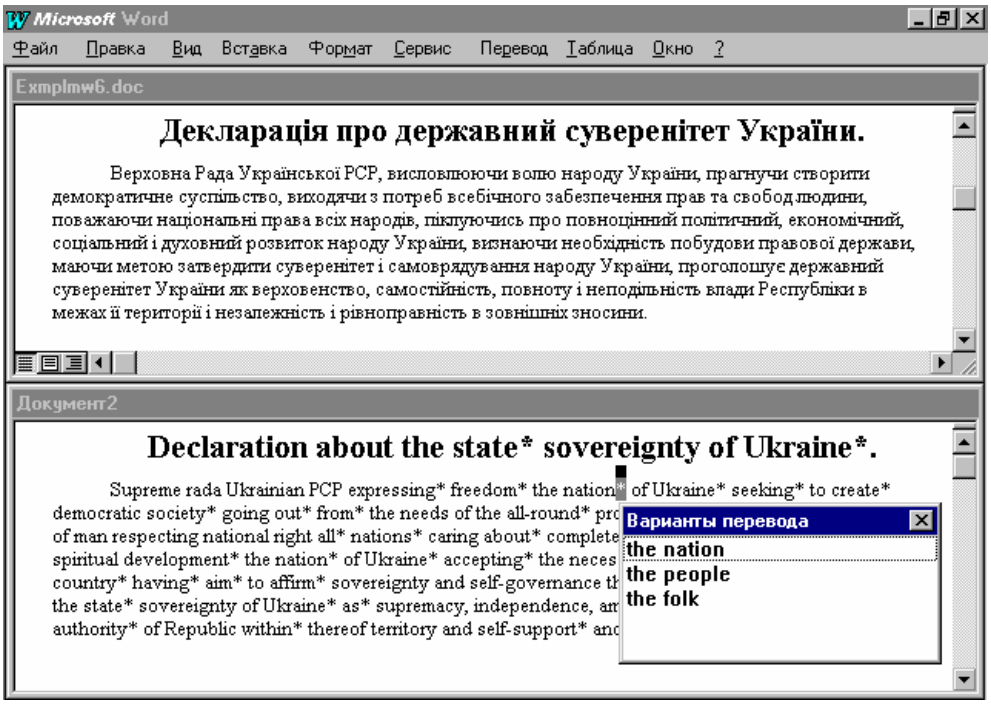

Figure 1: Translation by PARS/U

'New' words and phrases may be entered into the dictionary directly from the screen. The difference from the DOS version consists in the fact that the user marks the word/phrase to be entered, clicks the 'New word' button, and the word/phrase is written to the dictionary; i.e., unlike the DOS-versions, the principle is 'text first'. A further difference is that the Windows version allows not only separate words but also phrases to be entered into the dictionary directly from the text.

The user may also translate on-screen Help pages and texts of Internet WWW-pages written in HTML format. This is done via the Clipboard: the text portion to be translated is copied to

the Clipboard, the MT program is called up, and the target text appears in a separate window below the source text. Figure 2 illustrates the result of using PARS to translate an English help file into Russian. Figure 3 shows how an English HTML file has been translated.

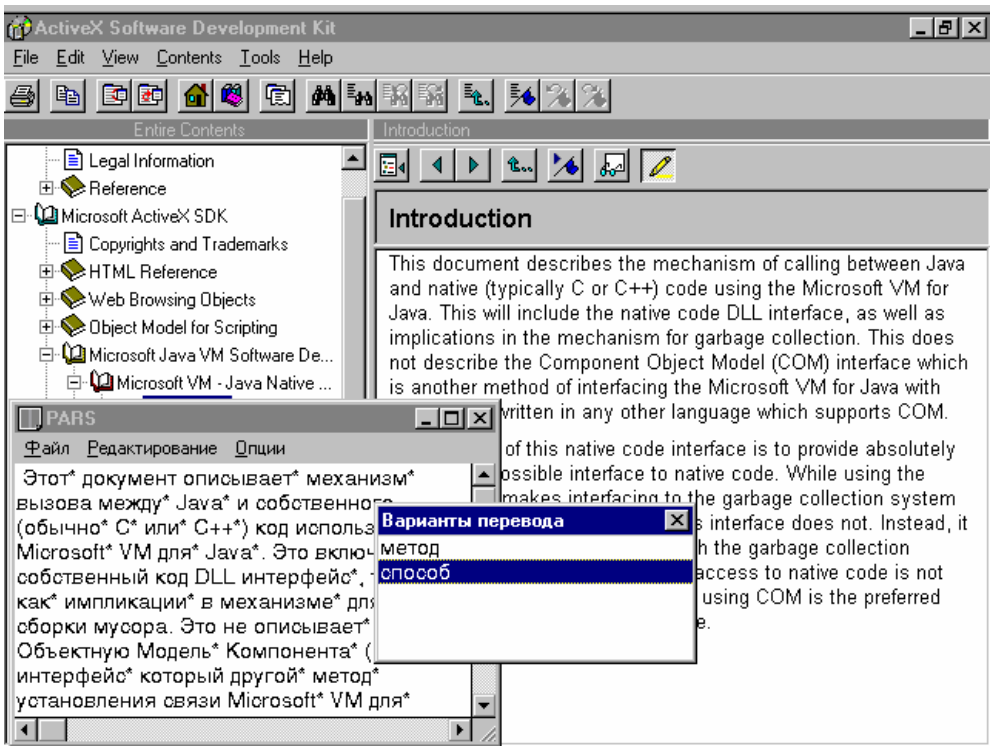

Figure 2: Translating a Screen Help File Screen using PARS (English-Russian)

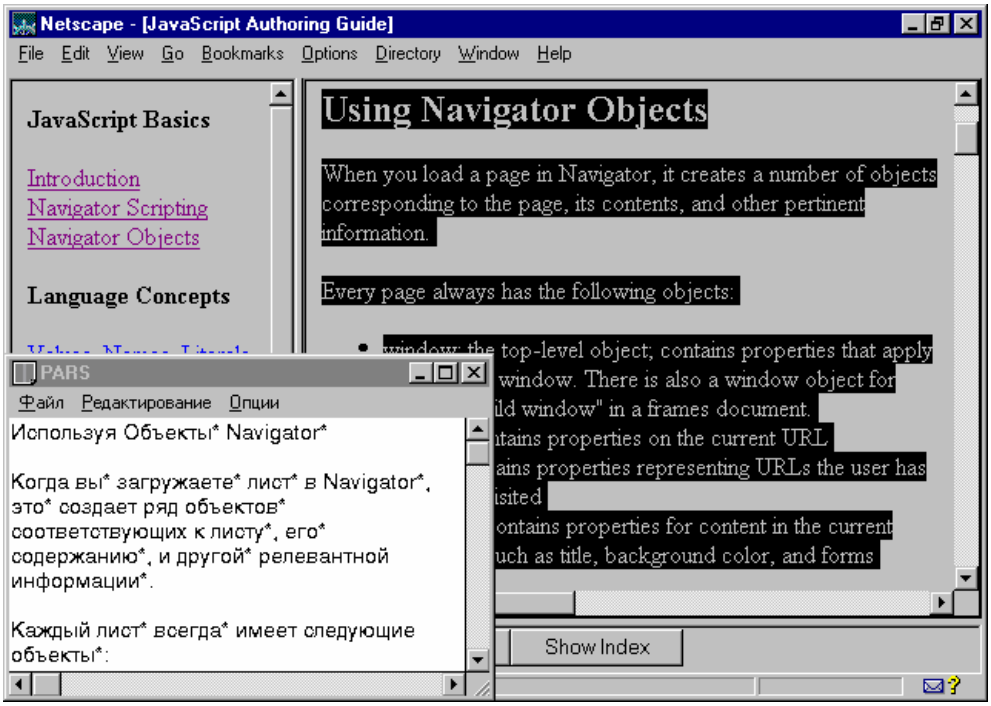

Figure 3: Translating an HTML File using PARS (English-Russian)

In each case the machine translation may be saved as a separate file. It should be noted that Lingvistica '93 and ETS are completing a new joint project which will enable the PARS program to be embedded directly into Netscape Navigator. Both DOS and Windows versions of PARS and RUMP run in stand-alone and network modes.

Translation: General Principles and Problems

As experience shows, it is rather hard to draw a demarcation line between the 'classical' translation strategies, direct and transfer-based. PARS, PARS/U, and RUMP have dozens of transfer rules, though it's hard to call them purely transfer-based systems. I prefer instead a different kind of terminology, distinguishing between the following two translation principles:

- FAT: **F**irst **a**nalyse, then **t**ranslate
- FTA: **F**irst **t**ranslate, then **a**nalyse

Our products are FTA-type systems, just like quite a number of very well-known PC-MT systems, such as those by Globalink. The system first translates the source text 'word by word' and 'phrase by phrase', and then tries to edit it according to the rules of the target language.

Let us be clear about one thing: if the source and target languages are not particularly close as, for example, Russian and Ukrainian, the output texts tend to differ markedly from those made by qualified human translators. When I hear or read that an MT system ensures '80-90% accuracy', I am inclined to consider such a statement a mere advertising trick. Yes, machine grammars are being constantly improved, but, being a professional language engineer, I can hardly imagine that computer programs will ever be able to compete with human beings.

The Capabilities of the Lingvistica '93 Systems

Despite the reservations about claims to translation quality made above, it is the case that RUMP indeed translates texts in such a way that they are 70-80%, sometimes even 90% ready for publication, the quality of the Russian-Ukrainian translation direction being somewhat higher than that of Ukrainian-Russian. The PARS and PARS/U MT systems are used for the following purposes:

- to give the user a general idea of the document's contents, for example, to browse large databases or 'scan' the text
- to create a draft translation for subsequent post-editing

The option of selecting translation variants essentially simplifies editing of the machine translation. This option also provides for the transliteration of proper names: for example, the Russian name Ivanov (in Cyrillic script) is not translated into English by PARS, but its transliteration (into Roman script) is suggested as a translation variant.

A number of users, including professional translators, maintain that 'it is very hard to edit machine translations in MS Word'. I will now explain what is meant by this statement.

The main disadvantage of using FTA-type programs to translate between languages, one of which, say, belongs to the Germanic group and the other to the Slavonic one, is that more often than not such programs fail to reflect the rules underlying word order in the the target language; it is up to the translator to correct the output. Getting the word order right in the target language requires complex transfer rules based not only on grammatical but also on semantic characteristics of the words; taking account of semantics in machine translation is a task for the

next generation of commercial MT systems. With regard to the problems of editing machine translations in MS Word, the only option available to the translator or post-editor is to rearrange blocks of text — often a tiresome process. That is why we are developing a Windows version of PenEdit. This will allow the user to transpose words very easily by means of an electronic pen or a digitizer.

MT systems supplied by Lingvistica ’93 may use up to four dictionaries in a single translation session; the user can set priorities for their application. When translating, the system looks up the word (phrase) in the dictionary which has the highest priority; if the item is not found there, the system consults the dictionary at the next level of priority, and so on. As it turns out this approach has both advantages and disadvantages. The drawbacks are as follows:

a) To begin with, PARS comprises a number of dictionaries, which requires linking more than four dictionaries in some translation sessions. For example, the following dictionaries may be used for translating aviation texts:

- general
- aviation
- aerospace
- mathematical (mathematical modelling in aircraft building)
- computer
- aviation medicine
- radioelectronics
- ground and space communications
- polytechnical.

b) Having found a word in one of the dictionaries, the system stops looking it up in the rest. This may produce an incorrect translation, simply because one and the same word may be present in different dictionaries and thus have different meanings.

c) Another problem consists in the difficulty of assigning the appropriate priorities to the dictionaries. For example, PARS translated an English medical text into Russian using the medical and general dictionaries in the indicated order of priorities. The result was that the English word ‘flow’ was translated as ‘menstruation’; ‘flow’ was in fact suggested as a translation variant and if the general dictionary had been assigned a higher priority, the translation would have been correct.

It seems to me that one of the most important criteria for evaluating a commercial MT system is its dictionary support subsystem: the easier it is to extend dictionaries supplied with the system as well as create user’s dictionaries, the better the system is in general.

The Lingvistica ’93 systems have a number of user options. These are listed below.

1) All dictionaries are fully bidirectional. For example, if the user enters an English word with its Ukrainian translation into a PARS/U dictionary, the system automatically sets the translation pair in the opposite direction, i.e. Ukrainian-English.

2) Any dictionary may be browsed and edited by the user.

3) It is very important to recognise that a word/phrase can have a practically unlimited number of translations. To take account of this any number of translation variants for a source item may be specified for the target text.

4) The form and layout of dictionary entries in Lingvistica ‘93 systems are reminiscent of those in traditional dictionaries. The difference is that, while in ‘paper’ dictionaries it is the

head word which is replaced with a tilde in a phrase (this word bearing the main sense of the word string), it is the first word that is considered as the head item in PARS, PARS/U, and RUMP dictionaries.

Figure 4 illustrates a dictionary entry in PARS/UNB (note that since Russian and Ukrainian are inflectional languages, word endings are separated from the stems with vertical lines).

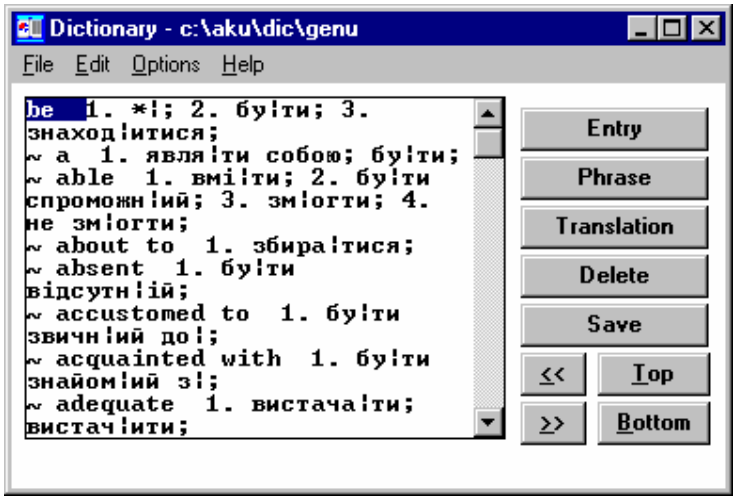

Figure 4: A Dictionary Entry in PARS/U

The user may use the one keystroke transposition option in the dictionary entry assigning a higher priority to the translation which is considered the most likely one for the subject area. For example, in the PARS general dictionary, the Russian word ‘obshestva’ has two English translations: ‘society’ and ‘company’. For translating socio-political texts, it is advisable to put the translation ‘society’ in the first position in the dictionary entry; the term ‘company’ is then included as a translation variant (marked with an asterisk character). For translating financial-legal texts, the order of equivalents is precisely the opposite.

5) These systems have a fully-automated indexing facilitity that tags Slavonic words as they are entered in the dictionary. The system automatically assigns grammatical features to such items, including the part of speech, declension, conjugation, and subclass characteristics (such as gender). If the program is uncertain how to index a word, it offers the user a choice between several options (for example, the declension of a Russian word).

The Lingvistica ’93 Dictionaries

An important feature of Lingvistica ’93 is that it uses primarily dictionaries produced by professional lexicographers. The lists of dictionaries supplied with the MT systems specify the names of their compilers.

PARS features a large spectrum of English-Russian-English specialist dictionaries, the subject areas being technology, business, medicine, space engineering, electronics, mathematics, chemistry, automobile building, etc. The total number of terms as of April 1997 is over 700,000 words and phrases in each language direction — English-Russian and Russian-English.

Such large dictionaries could never be compiled without collaboration between Lingvistica ’93 and ETS. Under the joint PARS+Polyglossum project, the entries of the world’s largest English-Russian base dictionary, Polyglossum, are semi-automatically converted into the PARS format. The procedure of semi-automatic processing consists of three stages.

1) The first stage is to import the Polyglossum dictionary into PARS.

2) Next, the Russian words of the new dictionary are encoded in batch mode according to the ‘coincidence principle’: the word acquires the same grammatical characteristics as in the PARS dictionary that was set as the prototype.

3) Finally, the dictionary officer looks through the dictionary entries and encodes the words that were not encoded by the batch mode program. In this case, the program uses the ‘analogy principle’: the word acquires the grammatical characteristics of similar words that have been entered into the other dictionaries.

Dictionaries are compiled very quickly, and the speed increases with each new dictionary, as the system has more and more encoded words to compare the new ones with.

If there is no Polyglossum dictionary for a particular subject area, the PARS dictionary is created by running a representative corpus of texts through the translation system with subsequent input of ‘new’ words and phrases into the dictionary.

However, we are fully aware that, in certain fields, dictionaries must be updated much more quickly and frequently than in others in order to take account of the most recent terminology. This is especially true, for instance, of an area such as telecommunications, which is ‘terminologically flexible’. The only way to achieve this is to cooperate with those companies that actually generate the new terminology. I would recommend such collaboration to all those who are concerned with translating technical documentation between Russian, Ukrainian, and English.

Translation Technology: PARS and Polyglossum in Tandem

Experience shows that in our case the most efficient way of translating from Russian into English and from English into Russian is to use PARS and Polyglossum to complement each other. In fact Polyglossum system has a flexible program for dictionary look-up, and the word entries in its dictionaries contain numerous explanations and commentaries. That is why Polyglossum is not only a source of new PARS dictionaries, but also serves for translating technical terms which PARS fails to translate or for choosing a more appropriate translation variant if the human translator post-editing the raw machine translation needs an explanation of a term.

The most recent example (February – March, 1997) of this technology in action was the translation from Russian into English of a collection of lectures on various branches of the aviation industry. The work was done by Lingvistica ’93 for Kharkov State Aviation University and Kharkov Aviation Plant. The total volume of texts to be translated amounted to several hundred pages. Translation was carried out using both PARS and Polyglossum. The dictionaries used and the number of terms in each are indicated below:

a) PARS:
- general (40,000)
- polytechnical (76,000)
- concise aviation dictionary (7,000)

- aerospace (60,000)
- mathematical (80,000)
- computers (20,000)

b) The Polyglossum polytechnical dictionary

The source texts were received as DOS and WinWord files, that is why both the DOS and Windows versions of PARS were used for translation. The source texts were first translated by PARS. After the machine translation of each document, 'new' terms (practically all of them were then found in the 300,000 term Polyglossum polytechnical dictionary), were entered into the corresponding PARS dictionaries, which substantially improved the quality of the translation of subsequent documents.

The machine translation underwent human post-editing, the purpose of which was to produce an informative, though possibly stylistically imperfect English text. Editing was carried out in two stages:

- primary editing: this was performed by two translators who have a quite good command of English grammar but do not specialize in translating texts on aviation;
- final editing: this involved checking of the terminology by an experienced translator of aviation texts.

In the context of this work, we tried to determine the efficiency of using the two systems, PARS and Polyglossum, at the stage of primary editing. The question that we put to the translators was: is it easier, and if so, to what extent, to edit the machine translation than to translate the text manually? The translators replied that it was three to four times easier to use machine translation. Using PARS and Polyglossum, each translator produced between twenty and thirty pages a day.

Another goal of this work was to improve the translation algorithm and determine the most frequent operations made by the translator when editing machine translations from Russian into English. This will permit A. Kazakov's group to fine-tune the PenEdit program.

How are the Systems Supplied?

Lingvistica '93 supplies the MT systems on the principles of either 'buy-and-go' or 'registered user'. In the latter case, the customer pays 350 Ukrainian grivnas (about $180) on an average for a single licence, or 500 grivnas ($260) for a networked version, after which, as a registered user, he/she gets an upgrade free of charge every four to six months for two years.

The 'buy-and-go' notion was suggested and implemented by Igor Fagradiants. ETS sells PARS and Polyglossum in Russia, and RUMP in Ukraine, on CD-ROMs: the prices are very low and reflect the low wage levels in the former Soviet Union. The average price is as low as $15 for a disk.

Michael Blekhman may be contacted at blekhman@lotus.kpi.kharkov.ua

The Ergo Parser Challenge

by

J. L. Morris

Derpartment of Economics, School of Social Sciences
University of Birmingham

Recently Phil Bralich and Derek Bickerton of the University of Hawaii have offered a challenge to users of the Internet. The challenge is to visit their website and try to 'outwit' their natural language parser by typing in a valid English language sentence that cannot be parsed or parsed correctly by the program.

The gauntlet thrown down by Phil Bralich and Derek Bickerton (B&B hereafter) poses a most interesting spectacle for those computer users such as myself who are keen to see ushered in a new era of computer-usage: an era in which natural language input and output becomes a realistic interface between man and machine. This after all represents almost the ultimate in user-friendliness.[1] Whilst single-sentence parsing is by no means all that is required to bring this new age to fruition, it would certainly represent an important milestone en route. But is it feasible that a machine could be programmed to deduce an identical parsing of a given isolated sentence to that produced by a human being?

In this review we will endeavour to explore how far Ergo is able to meet this goal and we will contrast its performance with two other contemporary parsers which are now widely available. Unable to afford a fast PROLOG compiler I have not been able to rank the comparative performance of Ergo and the various advanced PROLOG-based unification grammars which have been produced in the past few years.

In reviewing the Ergo parser we should realise that its aims are somewhat limited. Seeking only a single-parse for each input sentence it does not stand in competition with more powerful techniques of corpus-based probabilistic parsing. Nor does it appear to be based on an extensive semantic network such as WORDNET which is now being quite widely used, given its free availability over the Internet for academic research.

The two parsers which this reviewer sees as most likely to be in contention for the Ergo parser market segment would be Daniel Sleator's Link Grammar developed at Carnegie-Mellon University and the Good Language Software (GLSP) parser of Hristo Georgiev-Good.

The answer to the question posed in the second paragraph as to whether single-sentence parsing is possible hinges partly on just which human being the computer is supposed to be emulating. Each distinct socio-cultural background entertained for the potential human understander will impose its own semantic preferences on the interpretation placed on a text. But B&B claim only to be parsing single sentences rather than offering comprehension of running text. Does this lesser ambition render their goal any more readily attainable on a Personal Computer, given the current state of the art? Anaphor resolution is presumably not

[1] Given recent advances in the theory of cognition, one may surmise that eventually man and his/her personal computer will have a symbiotic relationship in which a shared consciousness is distributed between them — based on the idea of division of labout between mortal and silicon souls. In this mode of intimate working the partners perceive an occasional sensation of extra-sensory perception when their interlocutor is able to anticipate their conversational move.

even attempted, nor is analysis of metaphor — a crucial ingredient in a deep level understanding of running text or discourse. But how about context-sensitive parse attribution? Does Ergo have a strategy for approaching this crucial feature of what might be termed intelligent parsing.

The deductions associated with context-sensitive parsing are, of course, a major part of common-sense reasoning. Over the past decade, with the pioneering work of Lenat and Guha in their CYC system, great strides have been made in the codification of a knowledgebase of commonsense facts and rules, organised according to ontological groupings, with the domain-specific reasoning processes carefully associated with the appropriate categories. Two features stand out in this latter work. First, the good news is the crucial result that shallow inference is all that is required in commonsense reasoning to understand human communication. The bad news is that there is an enormous number of commonsense rules which have to be applied — and their identification, articulation and codification is an extremely labour-intensive process. As Lenat and Guha put it: 'there is no such thing as a free lunch.' In this reviewer's design of an Expert System for understanding economic text (UTILESE) the distinct contexts giving rise to alternative context-sensitive syntax, semantic or pragmatic features are referred to as WorldViewContextVignettes. Inheritance in a strongly hierarchical formulation allows some reduction in the size of the commonsense rule set.

The Real Role of the Challenge

Given the labour-intensive nature of eliciting commonsense reasoning identified by Lenat and Guha, it seems an excellent use of the Internet to try to harness the surfer's intellect in an essentially symbiotic way in order to perform the brute donkey-work which would furnish the commonsense knowledge base. In return the somewhat unsuspecting surfing punter is provided with entertainment, marvelling at the superiority of the human brain over the machine every time a new lacuna is probed. But is this what B&B are doing? It is tantalisingly difficult to deduce not only what parsing formalism they are using, as pointed out by Daniel Sleator, but also one is left completely in the dark as to the precise learning strategy being followed. The rather bland invitation to try again in a month's time, whenever one poses an unparsable sentence, suggests that the learning strategy probably corresponds to tweaking the lexicon rather than an automated context-sensitive language acquisitions facility.

The limitations manifested by single-sentence parsers such as Ergo are also clearly seen in another new exponent of the single parse philosophy, the Good Language Software (GLSP) parser of Hristo Georgiev-Good which has recently been advertised with a free one month's trial offer on the Internet. More will be said about GLSP later but for now we note that the problem with this approach is just that isolated sentences are quite often so ambiguous that any computer parser which merely explores a solitary parsing strategy is bound to be wrong a sufficient number of times to render it unsuitable for human computer interaction. Presumably Ergo will need to be able to recognise where such ambiguity occurs and to pose questions to the user to elicit further information to disambiguate the context.

Common-sense parsing, of the kind carried out by the CYC software, is able to make sentence-parsing a highly context-sensitive operation — capable of deducing from the preceding dialogue moves what the appropriate context is and adjusting the parsing algorithm accordingly.

The Bralich and Bickerton approach has a certain deja-vu connotation for this reviewer, having devised in the early 1980s a natural language parser which worked in the domain of

book-keeping (accounting) events. Called PACIOLESE, the Definite Clause Grammar-based parser written in PROLOG was able to deduce the appropriate parsing of the common kinds of transaction which characterise the recurrent activities of a business. Rudimentary language acquisition facilities were present in PACIOLESE, at least to the extent of what Carbonell called 'learning by being told'. New vocabulary could be deduced so long as the rest of the sentence within which it was embedded conformed to an easily recognisable form. Such a primitive learning mechanism is of course no longer acceptable as a mechanism for human computer interaction, given the enormous strides made in machine-learning over the past decade.

The one thing that the PACIOLESE program was not very good at doing was deducing an appropriate context when confronted with an isolated sentence totally unrelated to the realm of accounting-book-keeping. My colleague Professor Gambling devised the following test sentence: 'Blue Boy won the 3.30 at Kempton Park.' This sentence defeated the program to the glee, it has to be said, of most of the Accounting Department staff, who were anxious that their professional expertise was about to be dispensed from a robot at a tiny fraction of the cost of the going rate.

The 'Blue Boy' sentence evoked the ultimate lack of comprehension, signalled by the error message 'no known asset or transaction present in this sentence'. Strictly speaking, of course, the PACIOLESE response was completely accurate. Nevertheless this response was regarded as unacceptable and proved something of a turning point (the zenith) in the program's popularity.

Typing in sentences to the ERGO PARSER program produces that eerily familiar feeling to an ex-DCG programmer that the formalism at the other end of the network communication line, presumably in the University of Hawaii, is just not sophisticated enough to hold any kind of worthwhile conversation. This impression is strongly reinforced by the enjoinder to 'try again in about a month's time', whenever the challenger defeats the parser.

The Performance Characteristics of the Ergo Parser

On the evidence of a devastating critique of the Ergo Parser recently advanced in the Linguist Mailbase and elsewhere, notably by Daniel Sleator of Carnegie Mellon University, author of the freely available public-domain Link Grammar, one would be forgiven for thinking that the Ergo Parser is rather limited in the coverage which it offers of the English language, a limitation rigorously enforced partly by its maximum word count of twelve and its maximum character count of seventy-two. Sleator lists a substantial number of linguistic components that are not recognised by B&B. He also contrasts his Link Grammar multi-parse philosophy with the single-parse, take-it-or-leave-it methodology of B&B. In a rejoinder in the Linguist Mailbase a few weeks after Sleator's critique, B&B claimed that their latest version of Ergo could indeed now properly handle most of the troublesome cases that Sleator reported. Clearly the parser is in a state of flux given the 'on the job learning' which seems to drive its development.

This reviewer can confirm that most of the sentence types which Sleator claimed were not recognisable by the earlier Ergo are indeed now fixed. However, there is clearly an enormous number of possible ways in which Ergo could be challenged and found wanting.

I would deduce that it is not of great concern to the authors, Bralich and Bickerton, that arguably one of the world's greatest academic authorities on computational linguistics should

be able to outdo their parser. I would judge that they are attempting to create a parser which will cope with a limited kind of dialogue on the Internet; the cramping restriction on word count is probably designed to keep the number of distinguishable contexts down to a manageable proportion. As Sleator points out, Ergo will probably suffice for certain niche application areas such as Computer Games, inviting limited demands on the utterances from users.

The reader may gain some idea of the performance of Ergo from the following simple sentence, which, as Sleator tells us, was able to defeat an earlier Ergo: 'The bishop said he was coming.' The response is shown in the table on page 24.

As the reader will verify, although the basic parsing categories seem correct and useful, the subsequent transformations, particularly of tense, leave something to be desired in the way of conformity with the English language.

Conclusion

The crucial question is whether readers of the *Machine Translation Review* will find the Ergo Parser useful. In agreeing that Ergo may play a niche role in Internet applications requiring a limited dialogue such as Computer Games or Internet shopping, one is also recognising that the kinds of limited dialogue used in such contexts are precisely what is required in introductory tuition of ESL.

Whilst conceding that the final version of Ergo may indeed be useful, other readers of this Journal, in my estimation will probably find that the combination of the Sleator Link Grammar parser and associated resources taken in conjunction with the Good Language Software Parser and online dictionaries would provide a more comprehensive way of coping with their parsing needs.

Letting Ergo have the last word on this I tendered the following sentence somewhat tentatively: 'My guess is that you use a PROLOG Definite Clause Grammar.' This elicited the response: 'I can't parse that at this time, try again in about a month. Misspelled or Undefined Words: PROLOG.'

Appendix

Comparison of Sleator's Link Grammar and Ergo

The very valuable natural language parsing resource at the Carnegie Mellon University website maintained by Daniel Sleator gives not only the source code of his Link Grammar complete with a sizeable dictionary and banks of test sentences but also documentation and copies of published papers explaining the theory devised by the authors.

It is just conceivable that on simple sentences the Link Grammar could well be slower than Ergo. Though it has to be said that timing comparisons are not really possible since Ergo runs only in Hawaii and on a platform of unspecified description.

Nevertheless this reviewer has little hesitation in asserting that the Link Grammar will almost certainly be the more useful in an academic environment, given its multi-parse approach, its greater coverage of the language and, remarkably, its availability in source code form over the Internet. This last item means that the Link Grammar is therefore adaptable to one's own needs. As regards any possible adverse speed comparison, it is probably better to leave the

computer running all night on large tracts and know that at least something sensible will be there in the morning.

A Comparison of Ergo and the Good Language Software Parser

Based on its processing of the Sleator test data sets, I believe that the GLSP is more sophisticated than Ergo in its coverage of the English language. Unlike Ergo, GLSP does not suffer from the prohibitively small word count limit which precludes all but the most simplistically contrived sentences. Thus GLSP does at least make a reasonable attempt at parsing realistic sentences, including most of the Wall Street Journal tract.

Comparison of Sleator's Link Grammar and the Good Language Software Parser

One of the really useful resources which Daniel Sleator has made available at CMU is an extensive bank of test sentences which pose various levels of difficulty for the Link Grammar. In addition, a tract of text taken from the Wall Street Journal is offered for comparison purposes. On the version of the Link Grammar implemented by this reviewer in Microsoft C++ under the Windows 95 operating system both tracts worked very well.

The availability of a month's free trial offer of the Good Language Software Parser allowed a very significant test to be done. Running the Sleator test sentences and the Wall Street Journal extract through the GLSP, one quickly concludes that GLSP is incredibly fast and also very rarely on target with longer sentences! What is lacking is a context recognition mechanism — precisely the same omission as with Ergo.

For readers who know how to program, a fusion of the Link Grammar and the GLSP formalism could prove a useful compromise between completeness and speed.

References

Barlich and Bickerton, e-mail: derek@Hawaii.edu, bralich@Hawaii.edu; URL: http://www.ergo-ling.com

Sleator, Daniel, http://www.cs.cmu.edu/~sleator

Sager, Naomi (1981) *Natural Language Information Processing: a Computer Grammar of English*, Reading, Mass., London: Addison-Wesley

Georgiev-Good, Hristo, e-mail: Good Language Software <languagesoftware@access.ch>, URL: http://www.access.ch/languagesoftware/welcom.html

Morris, J. L. (1992) 'UTILESE: a Design for an Expert System for the Machine Understanding of Economic Texts', *Journal of Economic and Financial Computing*, Winter 1992

Lenat, D. B. and Gupta, R. V. (1989) *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*, Reading, Mass., London: Addison-Wesley

John Morris may be contacted at J.L.Morris@bham.ac.uk

Parser v4.00A parse: (3:0:0) 3 + 0 + 5 = 8 (180) 0.33 seconds

Parts of Speech
“The” is an indefinite article
“bishop” is a noun
“said” is a verb
“he” is a proper noun
“was” is a verb
“coming” is a verb

Parts of Sentence
“The bishop” is the subject of the verb “said”
“he was coming” is a direct object of the verb “said”
“he” is the subject of the verb “was coming”

Sentence Type
This is a statement.

Tense and Voice
Active Past Progressive

Simple/Compound/Complex
This sentence is simple

Statement to Question
Y/N Question
did The bishop say he was coming
WH Question
What said he was coming
What did The bishop say

Question to Statement
Question to Statement not necessary

Active to Passive
The bishop being said he was coming by

Passive to Active
Passive to Active not necessary

Change Tense
Simple Past – The bishop said he coming
Simple Present –The bishop says he coming
Simple Future – The bishop will say he coming
Present Modal – The bishop could say he coming
Past Progressive – The bishop was saying he coming
Present Progressive – The bishop is saying he coming
Future Progressive – The bishop will be saying he coming
Progressive Modal – The bishop could be saying he coming
Past Perfect – The bishop had said he coming
Present Perfect – The bishop has said he coming
Future Perfect – The bishop will have said he coming
Perfective Modal – The bishop could have said he coming
Past Perfect Progressive – The bishop had been saying he coming
Present Perfect Progressive – The bishop has been saying he coming
Future Perfect Progressive – The bishop will have been saying he coming
Present Perfective Modal – The bishop could have been saying he coming

The Telegraph and Systran Machine Translation Systems for Personal Computer: NLTSG Seminar

by

Derek Lewis

On Thursday evening, 24 April 1997, in King's College, London, the NLTSG held a seminar on Machine Translation. The speakers were Mr Craig Thomson and Mr Peter Angell of Endeavour Technologies, who demonstrated and answered questions on two PC-based MT systems, Telegraph and Systran. The focus was on English and French, with considerable time being devoted to analysing and comparing the performance of both systems on the basis of short test sentences supplied by the audience.

The Telegraph system

Developed by Globalink, Telegraph is designed for use on 32-bit computer architectures. It claims to be based on a 'new' transfer model for MT, with 'advanced linguistic capabilities' and an 'easy-to-use interface'. The minimum system requirements are as follows:

- 80486 processor running at 66 MHz
- Windows 95 or Windows NT 3.5
- 8 MB RAM (16 MB recommended)
- 24 MB hard disk space (28 MB for German)
- Mouse
- CD-ROM drive

Telegraph translates between English and Spanish, French, Italian, and German. It has import and export filters for most common wordprocessing packages (e.g. Microsoft Word, Corel WordPerfect, Lotus Ami Pro) and also for HTML and RTF. It can be integrated as a menu option within Word or WordPerfect, so that documents may be translated directly from within the wordprocessing application.

Telegraph is supplied with a core dictionary of over 200,000 headwords, to which individual words or entire whole dictionaries may be added. Ready-made subject dictionaries can be purchased separately or constructed by the user. A so-called 'stackable dictionary feature' enables subject-specific dictionaries to be specified for a particular translation. The system may be networked, with users able to work from and update a single set of dictionaries.

In contrast to the Power Translator (also by Globalink), Telegraph has evidently separated the user interface from the translation module. The latter, known as Barcelona, can be detached from the overall system and plugged into another application. This means, for instance, that a text produced for e-mailing can be automatically translated into another language prior to forwarding to the recipient. In addition to this, Telegraph offers a number of interesting features. Firstly, the user has access to the grammar or translation rules (via a 'rules editor') and can modify these in order to enhance the output. (Unfortunately the demonstrators were unable to present or explore this important aspect of the system.)

Secondly, the translation can be executed interactively. This means that the system presents alternative possible translations to the user, who can select these from an on-screen menu. In practice, however, this facility seemed on occasion to be disabled by the translation rules component (in the French sentence 'Les poutres étaient le batiment', for example, it would not allow the user to select and paste into the target text the alternative translation apparently provided for 'étaient', although such an option appeared to be available). Thirdly, the user can set Telegraph to perform a word scan for items not in the translation dictionaries before carrying out a full translation.

Systran

SYSTRAN PROfessional for Windows is a fairly new development. Until recently, Systran has been available only on large mainframe computers. The minimum system requirements for the PC version are as follows:

- 80486 processor running at 33 MHz
- Windows 3.x, Workgroups, Windows NT or Windows 95
- 16 MB RAM
- 15 MB hard disk space per language pair
- Asian languages require separate display drivers and text input methods.

Available language pairs for Windows are: English to and from French, Italian, German, Spanish, Portuguese, and Japanese; Russian and Chinese into English. For a non-Windows (DOS) environment Systran provides for: English into Danish, Finnish, Norwegian, Swedish and Arabic; German into Italian and Spanish; and French to and from German. English into Russian and Dutch is projected for early 1997. There are also pilot systems for English to and from Korean and for Serb-Croat into English. Systran comes ready supplied with a wide range of specialist subject dictionaries.

Both Telegraph and Systran were demonstrably fast in translation, although the speed at which Systran operated was quite remarkable: the suppliers claim speeds of about 150,000 words per hour (600-750 pages per hour, or five seconds per page) running on a Pentium PC. On the whole Systran was generally regarded by the demonstrators as the more powerful system, delivering consistently higher quality output.

Further details on Telegraph and Systran are available from:

Endeavour Technologies, Colette House, 234 Station Road, Addlestone, Surrey KT15 2PH, telephone 01932 827324 and fax 01392 827325

Derek Lewis

Conferences and Workshops

The following is a list of recent (i.e. since the last edition of the MTR) and forthcoming conferences and workshops. Telephone numbers and e-mail addresses are given where known (please check area telephone codes).

31 March–3 April 1997
5th Conference on Applied Natural Language Processing
Washington Marriott Hotel, Washington, D.C., USA
http://cs.nyu.edu/cs/projects/proteus/anlp97

21–22 May 1997
EAMT Workshop: Language Technology in Your Organisation?
University of Copenhagen, Denmark
Tel: +45 35329079, e-mail: sussi@cst.ku.dk http://www.lim.nl/eamt

17–18 June 1997
(SALT) Club Workshop on Evaluation in Speech and Language Technology
Halifax Hall, University of Sheffield, Sheffield, UK
Tel: +44 (0)114 222 1827, fax: +44 (0)114 222 1810/278 0972
E-mail: Gaizauskas@dcs.shef.ac.uk, http://www.dcs.shef.ac.uk

7–12 July 1997
ACL97: 35th Annual Meeting of the Association for Computational Linguistics,
ECACL97: 8th Conference of the European Chapter of the Association for Computational Linguistics
Universidad Nacional de Educacion a Distancia Madrid, Spain
Tel: +1 908 873 3898, fax: +1 908 873 0014, e-mail: acl@bellcore.com

11 July 1997
ACL97/EACL97: Workshop on Anaphora
As above
Tel: +44 (0)1902 322471, e-mail: r.mitkov@wlv.ac.uk

13–26 July 1997
EUROLAN97: Summer School in Corpus Linguistics
Iasi, Romania
http://www.infoiasi.ro/eurolan97.html

14–25 July 1997
ELSNET's 5th European Summer School on Language and Speech Communication
Katholieke Universiteit Leuven, Belgium
http://www.ccl.kuleuven.ac.be/ess97/ess97.html or http://www.elsnet.org

24–28 July 1997
TMI97: 7th International Conference on Theoretical and Methodological Issues in Machine Translation
Santa Fe, New Mexico, USA
http://crl.nmsu.edu/Events/TMI/

1–2 August 1997
EMNLP2: 2nd Conference on Empirical Methods in Natural Language Processing
Brown University, Providence, Rhode Island, USA
Tel: +1 607 255 9206, e-mail: cardie@cs.cornell.edu

11–22 August 1997
ESSLLI97: European Summer School in Logic, Language, and Information
Aix-en-Provence, France
Tel: +33 442 592073, fax: +33 442 595096 e-mail: esslli97@lpl.univ-aix.fr
http://www.lpl.univ-aix.fr/~esslli97

18–20 August 1997
WVLC5: NLP (SIGDAT): 5th Workshop on Very Large Corpora
Tsinghua University, Beijing, China, Hong Kong University of Science and Technology
e-mail: kwc@research.att.com, or: joez@lexis-nexis.com

22–24 August 1997
ROCLING X (1997): International Conference Research on Computational Linguistics
Academia Sinica, Taipei, Taiwan
Tel: +1 908 582 5296, fax: +1 908 582 3306

23–25 August 1997
IJCAI97: Workshop on Ontologies and Multilingual NLP
Nagoya, Japan
http://www.ijcai.org/ijcai-97/CfX/cfp.html

11–13 September 1997
CFP: 2nd International Conference 'Recent Advances in Natural Language Processing'
Tzigov Chark, Bulgaria
http://www.cogs.susx.ac.uk/lab/nlp/ranlp/97.html

17–20 September 1997
IWPT97: International Workshop on Parsing Technologies
Boston, MIT, USA
http://wwwseti.cs.utwente.nl/Docs/parlevink/sigparse/

16–21 September 1997
2nd Tbilisi Symposium on Language, Logic and Computation
Tbilisi, Georgia
Tel: +9 9532 382136, e-mail: chiko@contsys.acnet.ge

22–25 September 1997
ESCA: 5th European Conference on Speech Communication and Technology
Rhodes, Greece
Tel: +33 476 824336, fax: +33 476 824335, e-mail: esca@icp.grenet.fr
http://ophale.icp.grenet.fr/esca/esca.html

10–14 August 1998
COLING/ACL98: 17th International Conference on Computational Linguistics,
36th Annual Meeting of the Association for Computational Linguistics
University of Montreal, Quebec, Canada

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 (No.5, April '97) 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.