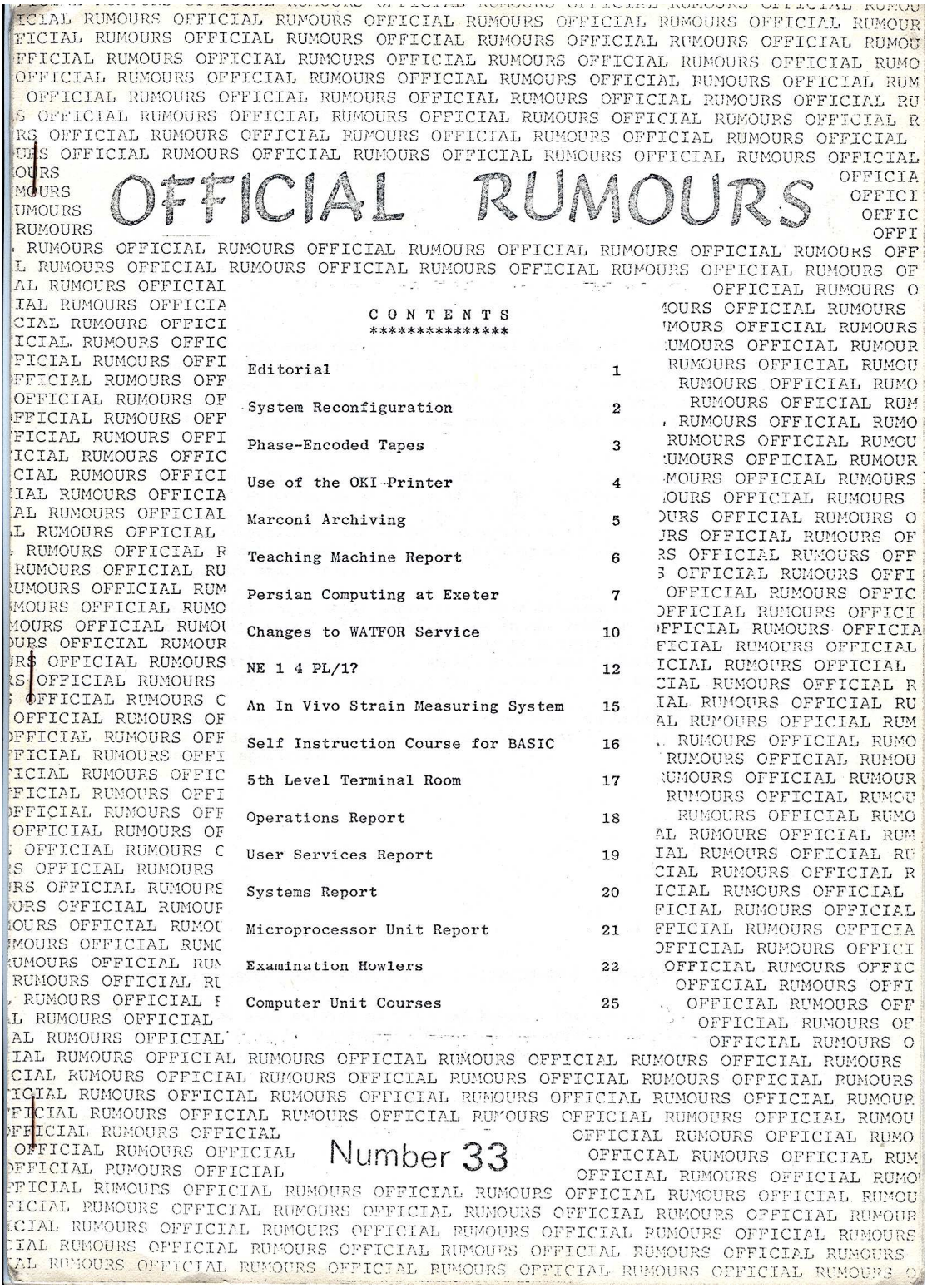
مقاله دربارهٔ نرم‌افزار واژه‌پرداز دوزبانهٔ فارسی و انگلیسی به‌نام Farsi System چاپ شده در Official Rumours انتشارات دانشگاه اکستر، انگلستان ، سال ۱۹۸۱.



OFFICIAL RUMOURS

CONTENTS
****

Number 33

OFFICIAL RUMOURS NUMBER 33 LENT TERM FEBRUARY 1981

UNIVERSITY OF EXETER COMPUTER UNIT NEWSLETTER

Editorial
****

We have made some changes to Official Rumours with this edition. In particular we have tried to include more news of the service, more reports of work in progress and more articles of general interest. I hope that you will find it better as well as bigger, and that it will be of use and interest to our growing number of users.

Your comments on the content and layout of the Newsletter would be very welcome, as would requests and suggestions for articles on particular topics. I would also be more than willing to publish letters to the Editor on subjects raised in articles, or on any subject connected with the Computer Unit, or the use of computers and microprocessors.

One article of special interest in this edition is concerned with the analysis and printing of texts in the Persian Language. The computer is being increasingly used by Humanities departments in universities throughout the world, and we are fortunate at Exeter to be able to offer such good facilities for this type of work.

We would welcome similar articles from users in other departments giving details of their use of the computer - especially in more unusual applications.

-

Jim Baker

-STOP PRESS-STOP PRESS-STOP PRESS-STOP PRESS-STOP PRESS-STOP PRESS-STOP PR

This is the last edition of Official Rumours which will be edited by Jim Baker, since he is leaving the ranks of the Applications Programmers to become 'Microprocessor Engineer' in the Microprocessor Unit.

Persian Computing at Exeter
****

The "Farsi" system was developed to produce Farsi (Persian) script as output from a number of linguistic text processing programs including the COCOA package which compiles concordances and frequency word counts of literary texts.

This system uses the Qume daisy wheel printer and accepts Persian data prepared through a simple and straightforward coding system. With very few minor alterations, the system may be utilized for the production of Arabic and Urdu script, and by changing the character wheel and using appropriate code set, it is even possible to produce other non-Roman alphabets, or special scientific and technical languages.

Special features and problems make Farsi different from standard Roman (English, French, etc.) scripts:

1. right to left direction for writing the characters, but
2. left to right writing order of numbers
3. attachment of most characters to the preceeding and following ones in the word (but not always!)
4. the variation of character sizes in length and width,
5. proportional spacing which is needed to handle this variation
6. different positional forms of characters (initial, medial, final and independent)
7. the very large character set
8. use of diacritics as short vowels, consonant repetition symbol and other phonetic features.
9. special forms for certain letter combinations
10. the importance of calligraphy in writing tradition of Persian and its legibility.

In the first instance, the 80 column card punch was used as the basis of coding system, since it is commonly available, even though the number of codes are very limited. But as an open system, the code set is capable of expansion to cover more special characters and function codes for different text analysis programs.

The present set contains all the persian characters, plus آ أ ؤ إ ئ ء combinations and ً ٌ ٍ َ ُ as diacritics , ten numerals, ؟ ! ، ؛ . as punctuation marks, - _ = # £ as linguistic analysis codes, and < > + (( )) as COCOA reserved characters.

Each Persian character is represented by one code character, the contextual position and the appropriate form is decided when the coded text is processed for output.

A hyphen (-) is used to distinguish internal spaces within words from spaces between words, and double brackets (( )) are used to enclose non-Farsi characters including English alphabet and any special characters or phonetic symbols which will be printed by the second print wheel on the Qume printer.

The data is initially prepared by relabelling the card punch keys with the Persian characters, using stick on labels. Since only one character is used to stand for each Persian character form, regardless of its position in the word, it is in fact easier to punch up the Persian than it would be to type it on a Farsi typewriter where it is necessary to choose different keys for each letter depending on its position. The method is however error prone, since the resulting code looks like a random collection of English letters and symbols which are difficult to correct in this form.

The programs necessary to convert this code into printable Farsi text are very complex, and would be extremely long and inefficient in most programming languages. Fortunately we have the SNOBOL string handling language available, which is ideal for this kind of application.

The conversion is done in two stages, firstly to aid debugging of the system, and secondly to make it more adaptable for other alphabets.

At the first stage each coded character is converted into a dual (character-figure) code, where the figures (0,1,2,3) indicate the position of the character in the word.

The second stage replaces each pair of codes with a series of Qume codes and characters, representing different components of a Farsi letter. One coded Farsi character may convert into as many as 10 actual codes which are sent to the Qume printer controlling spacing, positioning and over-printing.

For example the Farsi letter گ is shaped out of four symbols (ك ◌ٔ ◌ٔ ـ) and three small spaces. The mechanism of the Qume printer makes it possible to print from right to left automatically (by setting a negative horizontal character increment).

This is only an outline of the operation of the conversion program since all the problems detailed at the start of this article must be handled by the system.

The final printed text which is produced by this system is an accurate, well-formed and easily readable Farsi script of high standard.

Figure 1 shows a sample of text as initially produced on the card punch. Figure 2 gives a sample of the final Farsi script produced by the system.

S.M. Assi (Language Centre)
Jim Baker (Computer Unit)

Figure 1

```
AMA ZBAN-_HA_Y DYGR_Y =HSTND KH VA%_HA RA B(VRT SAXTMAN-_HA_IY
BA RVAB@ MTQABL BH-HM =MY-PYVNDND , M*LA: DR ZBAN 'RBY M(RY
RY$H-_I K-T-B (( K-T-B )) " =NV$TN " BA MYANVND (( A - A ))
TRKYB-_=MY-$VD , TA " K|T|B| " (( KATABA )) " ON MRD =NV$T "
T$KYL-_=$VD V BA MYANVND (( U - U )) TRKYB-_=MY-$VD TA " K`T`B "
(( KUTUB )) " KTAB_HA " B_DST-_=OYD . SAXTMAN_HA_Y NEVY GAHY
BA M@ABQH-_HA_Y $AQ V M$KL CVN JNSY;T DR ZBAN OLMANY V YA
M@ABQH-_HA_Y YKNVAXT CVN (( ILLORUM MAGNORUM VIRORUM )) DR
ZBAN LATYN M$X(-_=MY-$VND . DR B')Y ZBAN-_HA , TMAM JMLH ,
BA MFTAE NEVY YGANH-_AY M@ABQ BA (NF A$YAI_Y KH AZ_$AN GFTGV_=MY-$VD ,
=SAXTH-MY-$VD , MANND ZBAN " AYLAMBA " DR RVDZYA_Y $MALY ,
(( MONTO WAKOE MOKOLU OMOE WAKAKELA EKANA                          ))
" YGANH MRD BZRG AV BZRG_TR =AST AZ KVDK . "
(( KENTO KIAKOE KEKOLU KEMOE KIAMEKELA ENOMBA                      ))
" YGANH CYZ BZRG AV BZRG_TR =AST AZ XANH . "
```

Figure 2

اما زبان های دیگری هستند که واژها را بصورت ساختمان هائی
با روابط متقابل به هم می پیوندند ، مثلا در زبان عربی مصری
ریشه ك ت ب K-T-B " نوشتن " با میانوند A - A
ترکیب می شود ، تا " کَتَبَ " KATABA " آن مرد نوشت "
تشکیل شود و با میانوند U - U ترکیب می شود تا " کُتُب "
KUTUB " کتابها " بدست آید . ساختمانهای نحوی گاهی
با مطابقه های شاق و مشکل چون جنسیت در زبان آلمانی و یا
مطابقه های یکنواخت چون ILLORUM MAGNORUM VIRORUM در
زبان لاتین مشخص می شوند . در بعضی زبان ها ، تمام جمله ،
با مفتاح نحوی یگانه ای مطابق با صنف اشیائی که ازشان گفتگومی شود ،
ساخته می شود ، مانند زبان " ایلامیا " در رودزیای شمالی ،

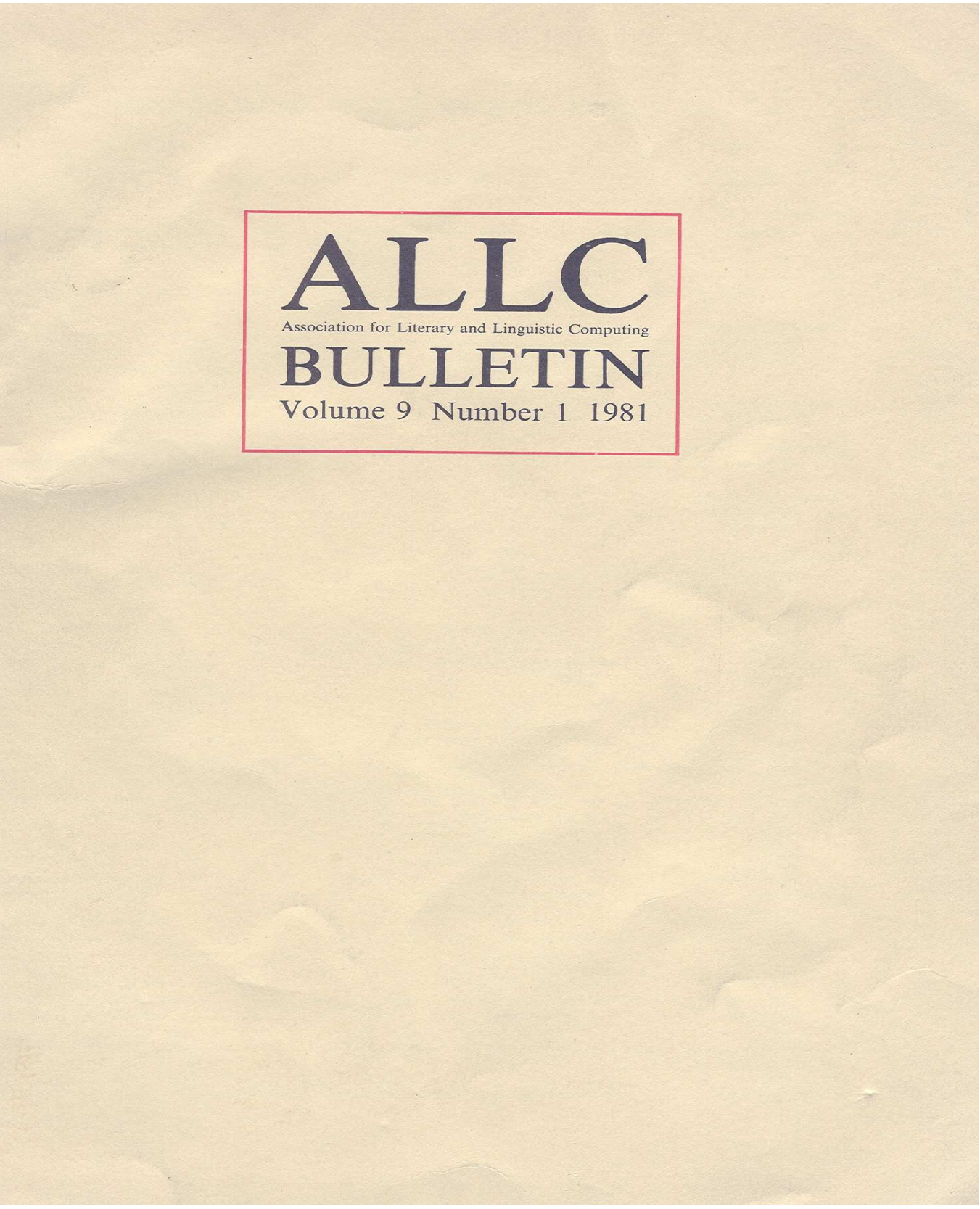
MONTO WAKOE MOKOLU OMOE WAKAKELA EKANA

" یگانه مرد بزرگ او بزرگتر است از کودك . "

KENTO KIAKOE KEKOLU KEMOE KIAMEKELA ENOMBA

" یگانه چیز بزرگ او بزرگتر است از خانه . "

گزارش دربارهٔ نرم‌افزار واژه‌پرداز دوزبانهٔ فارسی و انگلیسی به‌نام Farsi System چاپ شده در ALLC Bulletin انتشارات دانشگاه آکسفورد، انگلستان ، سال ۱۹۸۱.



ALLC Bulletin 1981
Volume 9 Number 1

CONTENTS

NEWS AND NOTES

Practice and Problems in Language Testing

Selected papers from the First International Language Testing Symposium of the Interuniversitare Sprachtestgruppe are now available as: Klein-Braley, C. and D.K. Stevenson, Editors, *Practice and Problems in Language Testing I: Proceedings of the First International Language Testing Symposium of the Interuniversitare Sprachtestgruppe held at the Bundessprachenamt, Hurth, 29-31 July 1979*, Orbis Linguisticus, Volume 1 (Frankfurt, Bern: Verlag Peter D. Lang, 1981), 222 pages, ISBN 3-8204-6203-1, cost DM 39, -/SFR 36, -. Full details are available from Christine Klein-Braley, University of Duisburg, FB 3, Lotharstrasse 65, D-4100 Duisburg 1, Germany.

A Guide to Information Sources

Gale Research Co. of Detroit have recently published the following guide: *Scientific and Technical Libraries in the Seventies: A Guide to Information Sources*, edited by Ellis Mount (Detroit: Gale Research Co., 1981). LC Card no. 80-25910, ISBN 0-8103-1491-6, $34.00, xvi + 157 pages.

British Standards Institution Publications

The British Standards Institution has recently published the following specifications, and has sent copies for the ALLC Archives at Cambridge.

BS 5887: 1980 Code of practice for testing of computer-based systems.

BS 5904: 1980 Computer programming language RTL/2.

BS 5905: 1980 Computer programming language CORAL 66.

Further information is available from the British Standards Institution, 101 Pentonville Road, London N1 9ND, UK.

'NIV Complete Concordance'

This concordance which claims to be the 'world's first completely computer-researched concordance' was described by three computer analysts and two biblical scholars at a news conference organised by Control Data Corporation and held at Multnomah School of the Bible, Portland, Oregon, USA on 8 June 1981. Processing was done on a CDC Cyber 175. The press release does not say how copies of the concordance may be obtained.

'Booke of Sir Thomas More'

Literary and linguistic computing scholars were invited to attend performances of *The Booke of Sir Thomas More* by the poor players at the Vandyck Theatre Bristol and the Young Vic, London during April and May 1981. This play has been the subject of a recent article in the *Observer* newspaper, 'Computer finds 'new' play by Shakespeare'.

Small Computer Program Index

Small Computer Program Index scans dozens of personal and small computer magazines of the UK and the USA (together with some other countries), to index thousands of printed program listings so that the user can find the precise reference to a listing on any subject, in any language or for any machine. Programs which have been listed in books are also included. The first issue of *Small Computer Program Index* will appear in January 1982. Each volume consists of 6 bi-monthly issues and an annual index in the last issue of each year. The 1982 subscription price is £15 ($35). Further information from ALLUM Books, 21 Beechcroft Road, Bushey, Watford, Herts WD2 2JV.

Current Problems in Textual Scholarship

The Society for Textual Scholarship sponsored a conference on this theme at the City University of New York Graduate Center, on 9-11 April 1981. Several of the papers included computer use in textual editing. Details from D. C. Greetham, CUNY Graduate Center, 33 West 42 Street, New York, NY 10036, USA.

Output With Persian Characters

A system called 'Farsi' has been developed at the University of Exeter, as part of a research project on English-Persian computer-aided lexicography, which produces Farsi (Persian) script on a 'Qume' Daisy Wheel printer. It accepts mixed English/Persian texts as input and prints them simultaneously. The character set contains all of the Farsi letters, diacritics and some special letter combinations. The input text is prepared in single codes and the shape of letters in connection with their contextual positions is decided by the system software, which also takes care of their proportional spacing and special combinations. Further information from S. M. Assi, The Language Centre, University of Exeter, Queen's Building, The Queen's Drive, Exeter EX4 4QH.

Output With Persian Characters

A system called 'Farsi' has been developed at the University of Exeter, as part of a research project on English-Persian computer-aided lexicography, which produces Farsi (Persian) script on a 'Qume' Daisy Wheel printer. It accepts mixed English/Persian texts as input and prints them simultaneously. The character set contains all of the Farsi letters, diacritics and some special letter combinations. The input text is prepared in single codes and the shape of letters in connection with their contextual positions is decided by the system software, which also takes care of their proportional spacing and special combinations. Further information from S. M. Assi, The Language Centre, University of Exeter, Queen's Building, The Queen's Drive, Exeter EX4 4QH.