[*Translating and the computer 10*. Proceedings of a conference … 10-11 November 1988, ed. Pamela Mayorcas (London: Aslib, 1990)]

International (tele)coms: a guide for the faint-hearted

Barry Mahon

European Information Market Development Group, Luxembourg

INTRODUCTION

Data communications, the use of telecommunications facilities to transfer data as opposed to voice, has been technically possible for many years. Starting from the rather simple telegraph services late in the last century it has developed rapidly in the 1980s to provide a sophisticated set of facilities for the movement of text and graphics over long distances.

Unfortunately and almost inevitably the use of these data communications has been restricted to enthusiasts or dedicated professionals, who have had the time and inclination to deal with the rather arcane world of the telecommunications engineer and the computer specialist. This paper is an attempt to demystify the world of telecommunications and to suggest some practical ways in which they may be used in the translation industry.

THE DEVELOPMENT AND PRESENT STATE OF THE ART

As mentioned above data communication started before the telephone, with the telegraph and the Morse code. It is ironic that the principles of the coding system invented by Samuel Morse are still being used today as the format for character transmission. Morse invented the process of using electrical signals generated by turning a switch on and off as a means of sending messages. Of course he had to connect the signals to an amplifier of sorts so that they could be heard and, more importantly, he had to invent a coding system to represent the letters. Morse's system was very simple using only combinations of long and short sounds – dots and dashes. In the same way, the transmission of characters depends on some agreed system or code for describing text in an unambiguous way before it can be transmitted. In effect, it is just another form of translation.

The code is needed to exploit the physical phenomenon whereby electrical signals can be transmitted from one place to another, either over a physical circuit or through the air, and can be detected at the other end. Morse's code was fine for its time but was superseded by the five-element code of Baudot, when the telex effectively replaced the telegraph and ultimately by the seven- or eight-element ASCII (American Standard Code for Information Interchange) code which is the basis of the transmission coding used in modern data communications. This very brief history of codes is necessary to lay the groundwork for an understanding of how and why text transmission can be effected reliably and efficiently.

One final word on coding: with the ASCII system, which is universally accepted by telecommunications administrations, it is possible to describe uniquely every character of every language in 10 so-called bits or binary digits. The slight snag is that not all telecommunications authorities recognise quite the same sequence within the 10, but I will come back to that problem later when I discuss accents and diacritical marks. When I use the word 'character' in the ASCII context I do not include pictograms although they can also be transmitted.

For many years the telex was the only universally available transmission service for text. It has a limited (agreed) character set and a relatively low speed of transmission – approximately 50 characters per second. The telex machine was, and still is, an electro-mechanical device which limits the possibilities for upgrading the technology: its universality is probably its only long-term value and the main reason why it has endured so long.

In recent years the facsimile or fax machine has become more popular as a means of data transmission especially for graphics. The technology has been available for many years but has been significantly improved recently. As an aside, it is interesting to note that the upsurge in the use of facsimile transmission has coincided with the needs of one nation in particular, Japan with its pictogram-based alphabet, to improve and speed up text transmission. In many circumstances, fax has replaced telex as a text transmission medium although its primary value is for transmitting documents containing graphics. This is despite the fact that telex is very often the cheaper alternative; it is probably because it is easier to transmit a fax where a paper copy of the text already exists, and today this paper copy is more than likely to have come from a word processor.

In parallel with these developments, the computing world was beginning to create the means for transmitting and receiving data originated in computers. Here again the technology was available before the computer itself was developed but it needed the invention of the ubiquitous chip to provide the means whereby data could be transmitted and received at speeds at which it became worthwhile to use another, more distant, computer to perform tasks or, to use the accepted phraseology, to share resources between computers.

The early manifestations of this capability were the airline booking systems which gave individual travel agents access to the central computer of an airline, effectively in dialogue mode, and allowed them to check the availability of flights and seats instantaneously. This was quickly followed by so-called time-sharing services which enabled large processing tasks, for example, calculating the pay of a large workforce, to be uploaded to a big computer centre and having the results – the pay slips – downloaded, or transmitted back, to the originator. Such was the power of these computers that they could appear to be doing a number of tasks at the same time.

Inevitably these resource-sharing activities became more widely distributed and needed suitable transmission media. The most easily available medium was the telephone network which was accordingly adopted and adapted to the customers' needs. The first need was to change the characteristics of the telephone network from that of a transmitter of voice signals to that of a transmitter of computer data signals. These signals, or bits, are digital in character whereas the telephone network is, or was exclusively at that time, analogue in character, i.e. it imitated and transmitted the frequency changes of the human voice. The computer by contrast spoke in only two frequencies, i.e. the bit was electronically on or off. The solution to this incompatibility was the modem (modulator-demodulator) which is attached between the computer and the telephone lines to translate – if that is not a misuse of the word – the computer's digital signals into the analogue frequencies required by the telephone network.

Progressively therefore, computers became more and more loquacious and needed more and more telephone lines to transmit their data. The requirements for speed meant that they could not use the ordinary lines that we use for telephone calls but needed dedicated circuits which had to be specially tuned to give high reliability. The obvious solution, as demand increased, was to create special telecommunications networks for computers to talk to one another and avoid using up telephone circuits.

Thus was born the data network, the telecommunications network that does not speak – it whistles. NASA in the United States was one of the first owners of a data network but was quickly followed by the airlines, the banks and big businesses and, of course, the military. In fact the military's requirement for bomb-proof data communications led to the development of one of the modern techniques of data transmission –

packet switching, but a discussion of this is unnecessary at this stage.

In Europe, the first dedicated transmission networks were also used by the airlines and the military. They were private in that they could only be used by their owners. In the United States such restrictions did not apply and before long data networks for pay-as-you-go clients who only wanted limited facilities for short periods began to appear. In Europe we looked on enviously because there were many who needed data communication facilities but who could not afford or justify a dedicated network. As we waited for our telecommunications suppliers (the PTTs) to respond to our needs, the European Communities (EC) took a hand and funded Euronet, a dedicated data network for the information industry. The PTTs, galvanised into action by Euronet, finally responded by providing public data networks which could be accessed by users on a 'need' basis. These facilities coincided with the wider distribution of computing power through the advent of the personal computer (PC).

The public data networks are interconnected so that it is possible to make a data communication connection to a computer anywhere in the world, in much the same way as can be done with the telephone. Every developed country and quite a lot of the developing countries have a national data network. This overcomes the compatibility problem. The fact that, for example, an IBM PC and an Apple Macintosh are not technically the same does not prevent them from communicating via a public data network. Controversy over restrictive practices being exercised by airlines using Amadeus and Galileo, two technically different reservations systems, has arisen, in part, because of arguments over the use of proprietary technologies for the data transmission, and not because of the public data networks.

APPLICATIONS OF DATA COMMUNICATIONS FACILITIES FOR TRANSLATORS

The first and most obvious application is for transmitting and receiving texts to and from machine translation facilities. It is accepted that effective machine translation requires large computer resources and these are economically viable only when many users share the facilities. That is not to say that today's facilities are being used in quite that way, but it is a basic requirement for the future development of machine translation resources that large centres be created, which are connected to data networks and which can handle many requests.

The second application is for transmitting and receiving texts for (non-machine) translation. This is linked to the use of word processing equipment for creating texts. Thus a translator can receive a text which has been prepared on a word processor, translate it either on or off screen and finally transmit the translation, via the word processor, to the client.

A variation on this process is that a translation bureau receives texts from clients and forwards them to translators who then return the translated text to the bureau, which adds the invoice, before retransmitting it to the client.

Another application of data transmission is in the use of large terminology banks such as Eurodicautom. Here the databank is held on a central computer and users call in via data transmission networks to query files and look for term equivalences. The Eurodicautom databank is operated in just such a fashion by Echo (the European Commission Host Organisation). Eurodicautom provides a service in all EC official languages but also includes terms which originate from other languages where this has a bearing on the meanings. The databank does not merely provide one-to-one equivalences; it also provides contextual data to aid the translator. Obviously such a facility is expensive to create and maintain and cannot be easily duplicated. Therefore its use as a central facility with data communications is more economical. Another similar use is for access to subject databases as an aid to identifying subject-related terminology in physics or medicine, for example. Here the translator can access a database on a particular subject and look for particular words or phrases in the context of material on a specific topic. This facility is often useful in newer areas of technology where terminology is not yet stabilised.

Thus we can see that there are a number of areas where data communications facilities can be used as a significant productivity aid in the translation industry. It remains now to discuss exactly how translators can have access to these facilities.

THE THEORY BECOMES PRACTICE

The first thing to be said is that entering the world of data communications is not an exercise for the faint-hearted. While this paper is designed to demystify telecoms and is dedicated to the faint-hearted, there is no point in pretending that it will all work first time and forever: it will not. My view is that things are getting better and that it is possible to do things via data communications with a little patience, some knowledge and some luck.

Hardware

Today there is little point in considering any terminal device (here, terminal means equipment used for establishing a connection to remote services, located at the user’s end) other than the personal computer. This may simply be a machine dedicated to word processing built around a computer chip, such as some of the Amstrad (also known as Schneider) models, or a full-blown PC made by IBM or one of the clones, or an

Apple. All of these machines should have a serial port which allows you to plug into the modem referred to earlier.

Despite the fact that there are dedicated data networks we still need to access them from our office or our home. Here, we normally have just the telephone so the modem is still needed to straighten out the computer signals until they reach the data network. There has been much discussion on the reliability of these data networks following tests done by organisations such as Eusidic (the European Association of Information Services), which indicate that approximately one call in three to or on a data network fails. However, it represents the state of the present art and is more often than not due to overloading either of the local telephone system before traffic reaches the data network or in the network itself. It is not quite as bad as it sounds; the situation is much worse in some countries than in others, and even varies from place to place within a country. Many users experience almost no difficulty.

Software

I will come back to the modem and its supply and installation later; I want to turn now to the software, the magic that turns the PC into a (tele)communications device. Nearly all PCs sold today come equipped, as explained above, with a serial port (also described as a modem port, a communications port or an RS232 port). This is the device that enables a PC to connect to a modem and thereby to the world of data communications. What is needed to actually make the connection is a piece of software which can address, as the jargon has it, the serial port. Some machines come with such software built into the basic facilities. There are also the integrated packages, such as Symphony or Open Access which provide a complete range of facilities including communications. Then there are the dedicated software products (programs) for communications such as Xtalk, Procomm, PC Term or MacTerm. My preference is for the dedicated product since I believe that programs written for a specific task work best, by comparison with the ‘do-everything’ packages. Whatever package you use, it should allow you to do the following:

- change speed, parity, number of stop and start bits and number of databits for transmitted or received data
- upload and download files (texts) in plain ASCII format as well as transfer data in protected form using protocols such as Kermit and Xmodem
- suspend the communication mode in order to carry out local tasks (e.g. word processing)
- transfer to and from directories other than that containing the communications program.

It is not too serious if some or all of the above terminology is unfamiliar as most of it is not necessary for routine communications. However, the capabilities mentioned are important for certain tasks, for example, for moving texts containing accented characters.

The problem of extended character set transmission

Many people who have a personal computer or a word processor come across the problem of representing accented characters at the level of keyboards – for inputting the characters, of screens – for viewing them, and of printers – for printing them. There are no accented characters in the English language and therefore it was calculated that 7 bits (binary digits) could be used to represent all the characters and commonplace punctuation marks. In order for ASCII to become universal or rather, to enable US computer manufacturers to sell to people who use 'foreign' character sets, it was necessary to extend ASCII to 8-bit representation in order to make room for all the accented characters. Unfortunately, before this was agreed, some European manufacturers had gone ahead and agreed on a 7-bit representation of the most used accented characters. As Dr Jackson has said in his paper, this has resulted in a form of anarchy which seriously inhibits data transmission to this day.

I mentioned at the beginning that the ASCII coding system is used to represent characters for data transmission, with 10 bits being used to carry the character. I also mentioned that not every PTT deals with it in exactly the same way. Public data networks expect 7 bits to be dedicated to the description of the characters that are being transmitted, with the other 3 bits being used for control purposes. As explained above, some accented characters require 8 bits to be transmitted in their entirety so some problems do arise when transmitting these over the public data network. Just as a matter of interest, this problem would not arise if the data was being transmitted over the ordinary telephone network but this would require that the equipment conditions at either end were exactly the same for a successful transmission. It will not be a problem when we get the so-called ISDN (integrated services digital network) but do not hold your breath! The present public data networks do permit different equipments to speak to one another and the way around the 7-bit/8-bit problem is to use protocols such as Kermit or Xmodem. These are not coding systems but pieces of software which, when loaded at both ends of a communication link, enable you to mask the data from the network and, as a bonus, check that what was sent actually arrived. This is done by so-called error-checking routines, the Kermit or Xmodem at either end, which speak to one another via the data network and carry the data between them. For error-free transmission of accented texts, it is essential to use these error-checking protocols.

Using data networks

With the software installed it is now only necessary to obtain access to the national data network; PSS in the United Kingdom; Transpac in France; Tymenet, Telenet or others in the United States. Application is usually made through a local office of the telecommunications supplier. In several European countries modems can only be obtained from the telecommunications authority, although this is changing. One important point about modems is not to buy a cheap one from a non-specialist outlet. Very often there is no support available and if it does not work you are on your own. Often, purchasing or leasing it from the telecommunications authorities is the best option; that way you can refuse to pay the bills if it does not work! There are many organisations such as Aslib who provide information on the mechanics of obtaining the necessary permissions and equipment and give information on user groups where more specific help may be available. The Annex lists centres in a number of European countries where advice and assistance can be obtained. With all the pieces in place it should then be possible to connect to whatever facilities you need. Finally, I shall describe one or two specific possibilities.

SCENARIOS FOR TRANSLATORS

I am acquainted with an organisation called GeoMail, an association of what is called 'added value message handling services'. This is a general description for a group of companies who provide a data communications-based service for those who wish to exchange messages. A message can be anything from a one-line telex to a ten-page document (for translation). The GeoMail association provides its customers with a one-stop shop for all data messaging.

In practice, a user can connect to the local member of the association, for example in France, Germany, Italy, the United Kingdom or the United States and all other services are available through this one point of entry. All services provide electronic mail (e-mail), a system which allows users to put a message in the 'postbox' of another user, read messages which have been delivered to their own box and send messages to a mailing list. All services provide access to the telex network, so messages can be sent to a telex subscriber. Some provide individual telex numbers for users so that telexes can be received even though the user does not have a telex machine. All services provide access to the ordinary telephone network for sending facsimiles. Users can also send and receive texts using the Xmodem protocol, which preserves the structure of the document as it crosses the data network. Such a network of services could provide obvious benefits for a group of translators. Let me put to you the following scenario, all of which is possible today.

An association of freelance translators offers a single point of contact for clients: clients can submit texts by post, telex, fax, or through e-mail. The coordinator redistributes the texts to individual translators via the e-mail facility, after having (in the case of material arriving by post or fax) keyed it into a word processor. The translator downloads the text into a word processor, performs the translation and then uploads the result back to the co-ordinator, or direct to the client. The returned text can, if necessary, be sent as an e-mail message, a telex or a fax. It is even possible to price translations on the basis of the turnaround speed associated with a particular transmission medium.

All of the above is possible as long as each member of the association has a communicating PC and a modem. They need not necessarily have the same type of PC and they need not necessarily be in the same country or even on the same continent.

In such a telecommunication based association a number of additional facilities can be made available to members. They can have computer-based bulletin boards or conference systems for posting notices of general interest, for discussing with colleagues any particular problems that have arisen with a text, or the translation of a particularly interesting word or phrase. They can, of course, use the telecommunications networks to submit and receive texts for machine translation. They can access terminology databases or subject databases for more specific information needs. In short, a whole world of facilities can be available from a local terminal, giving translators great flexibility in their working and domestic lives.

CONCLUSION

As I explained earlier, the world of data communications is not for the faint-hearted. However, the range of computerised aids which is becoming available at the other end of the data or telephone line is such that it is becoming interesting, not to say essential, for the growing army of homeworkers, to take their courage in their hands and make the plunge. I think translators may be delighted with the result.

AUTHOR

Barry Mahon, European Information Market Development Group, PO Box 262, L-2012 Luxembourg (GD).

NOTES

GeoMail, Infotap SA, BP262, L-2012 Luxembourg (GD).

Eusidic Report available from Eusidic (European Association of Information Services), 1st floor offices, 9/9a High Street, Calne, Wiltshire (attention Harry Collier). Tel: (0249) 816407.

British Telecom: *International Data Services — a Guide for Business Personal Computer Users*. Write to: British Telecom International, Holborn Centre, 120 High Holborn, London EC1 2TE. Tel: (01) 492 2000.

Department of Trade and Industry, 1 Victoria Street, London SW1H 0ET. Tel. (01)215 7877.

ANNEX

Centres providing assistance to data communications users

Belgium: CNDST, Boulevard de l'Empereur 4, B-1000 Brussels.
Tel: (2) 519 56 44.
Denmark: Sigurdsgade 41, 2200 Copenhagen N.
Tel: (31) 81 66 66.
France: Information Centre, Mr M. Cordelier, ANRT-Europe, 16 avenue Bugeaud, F-75116 Paris.
Tel (1) 47 04 47 57.
Germany: GID, Lyoner Strasse 44-48, D-6000 Frankfurt 71.
Tel: (6)1 16 68 71.
Greece: National Doc Centre, 48 Vas. Konstatinou Av, 11635 Athens.
Tel: (01) 724 0310.
Ireland: Eolas Glasnevin, Dublin 9.
Tel: (1)370 101.
Italy: DIANE Centre, via Cesare De Lollis 12, 00100 Rome.
Tel: (6) 49 52 351.
Luxembourg: PTT for Luxpac, M. R. Wennmacher, Administration des P et T, Division des Télécommunications, Services des Abonnés au Téléphone, 2999 Luxembourg.
Tel: 4991722.
The Netherlands: COBIDOC, Postbus 16601, 1001 RC Amsterdam.
Tel: (20) 22 39 55.
Norway: DIANE Centre, Tekniske Universitetsbibliotek, Hogskoleringen 1 Trondheim.
Tel: (7) 59 5120.
Portugal: CDCT, Av Prof Gama Pinto 2, P-1699 Lisbon.
Tel: (1)762 891.
Spain: CIDC, Calabria 168, 08015 Barcelona.
Tel: (3) 425 1799.
Sweden: DIANE Centre, IDC-KTHB, S-10044 Stockholm.
Tel: (8) 787 89 60.
United Kingdom: Aslib, Information House, 20-24 Old Street, London EC1V 9AP.
Tel: (01) 253 4488.

AUTHOR

Barry Mahon, European Information Market Development Group, 177 Route d’Esch, L-1471 Luxembourg.