

Number Language and Information Retrieval by Walter Goshawke

1 The word processing revolution

We are all aware of the current revolution in word processing, replacing the typewriter by the computer terminal, but not everyone realises what repercussions there will be in wider fields, one of which is the field of information retrieval, and how necessary it is to keep a close watch on developments. A letter typed in London can be instantly printed in Manchester. The Post Offices of the world are planning a similar service on a world wide basis. Already you can dial three quarters of the world's telephone subscribers from your office. Eventually you will be able to send and receive letters in the same way. All the necessary communications expertise is available, but there are human problems, one of which is comprehensibility. It's quite easy if both you and your correspondent speak English, but what if one of you cannot, but only speaks Russian or Arabic? I do not know what the authorities will do about this, but to me there is only one solution, machine translation. Furthermore, because of the enormous variety of different alphabets in the world, it must be machine translation by a certain method. All messages must first be translated by computer into Number Language, an unambiguous intermediate language consisting entirely of numerals. The Number Language message is then sent anywhere in the world and on arrival is translated by computer into the local language. A method is already available for use in business correspondence, and will work successfully if you word your correspondence carefully and avoid ambiguities. This is such an easy and necessary development that one feels that it is certain to take place. But human institutions are unpredictable. If it does not, the new network will be so crippled as to lose 90 per cent of its possible world wide potential. This would be a grievous loss to the business communities of the world and to all of us.

2 The Number Language dictionary

Let us examine the situation, however, if the suggested method is adopted. As soon as machine translation by Number Language has been established for business correspondence, it will quickly spread to technological and scientific translation. Many Number Language texts will be created and some will find their way into databases for instant translation when required into any language in the world. But before this can happen we must have technical and scientific dictionaries, and there are many branches of science and technology. It is at this stage, or earlier, that information retrieval specialists should take an interest. The creation of the Number Language dictionary is an enormous task, and great care is necessary in planning it if we are to maximise its usefulness for information retrieval and other purposes. As far as machine translation is concerned, any number can represent any word, and there is a danger that in the early stages numbers will be allotted in a haphazard fashion, and become so thoroughly established that users will be unwilling to change them to conform with an overall planned system.

All Number Language words have 10 digits, thus providing plenty of scope for present and future vocabulary. A scheme should be devised whereby the first four digits indicate which branch of science (or whatever) the word belongs to. A similar scheme is in operation for telephone numbers, the earlier digits having a geographical significance and the later ones being arbitrary. Librarians also have a systematic numbering method. If the Number Language dictionary had this feature, it would be possible to scrutinise any Number Language text by computer to ascertain its subject matter and also to identify words which receive intensive treatment. These details could be recorded in an index. This method would also have the advantage that whenever new entries were needed in the dictionary due to progress in research, changes in fashion, or any other reason, practitioners in the field concerned could be left to allot numbers in their appropriate series

without consulting the whole body of users, but would of course be required to report any new additions at prescribed intervals. If ever the supply of words becomes inadequate, double words (20 digits) could be used, giving many millions more. I visualise the Number Language dictionary eventually including every fact, idea and fancy ever mentioned in print in any part of the world, all thoroughly classified. I also visualise a huge collection of Number Language texts, efficiently indexed and growing daily, any document from which could be summoned in minutes to any terminal in the world, translated into the local language.

3 Is it possible?

But is all this possible? Is machine translation capable of performing this task? I am satisfied that it is. Machine translation cannot be successful without pre-editing. There are several reasons for this, perhaps the worst difficulty being ambiguities. If a text has an ambiguity, it sometimes happens that nobody on earth can resolve it except the author. Pre-editing is a nuisance, but it is a necessity. My own belief is that in due course writers will do their own pre-editing as they write their work, and when they present it, it will already be translatable by computer. How easy will it be for writers to get accustomed to writing in a style suitable for machine translation? I think it will be quite easy. Before very long, computer terminals for foreign correspondence will be in daily use in offices, universities, schools and some homes. In due course many people will have had experience in writing for machine translation continuously since their schooldays. It will be second nature to them. While there are these well-known difficulties in translating from spoken languages into Number Language, there are no such difficulties in the opposite direction. Number Language texts are unambiguous. Number Language has been specially designed for machine translation and its structure gives every possible help to the programmer. Number Language texts are ready for translation at all times. There is also the need for dictionaries. The creation of the dictionaries is an enormous task, and a highly skilled one, but all the techniques are known and it would only be a question of time before dictionaries in the major languages would be available for many branches of science and technology. Once established, the dictionaries would need constant revision to cope with new developments. The creation of the whole system would be very expensive, but it could be undertaken piecemeal. At every stage it would be usable, and would rapidly repay the initial cost. The cost of maintenance would be trifling compared with the enormous advantages of having the system in daily use.

4 What should be done?

I hope I have convinced my readers that the revolution in word processing is something in which they should be deeply concerned. It is up to them to do their best to see that any systems established are planned in such a way as to suit them, and thus be of the maximum possible use to mankind. There are a number of dangers. One is that local groups of countries may establish systems of communication between themselves based exclusively on the local languages and become so entrenched that it will be impossible without enormous expense to introduce a world wide system at a later stage. Another danger is that certain technologies, well in advance of others, will create Number Language dictionaries in a manner which will make it impossible for other technologies to dovetail with them in a logical manner. How are we to ensure that we get a planned system which will provide the maximum advantages for everybody? I suggest as a first step that all information retrieval specialists should make a thorough study of Number Language and do all in their power to persuade educational establishments to provide courses and engage in research in Number Language theory. Only by having a large, informed and vocal body of opinion dedicated to the aim of making machine translation as familiar a feature in our lives as the telephone or television will it be possible to influence the course of events in the desired way,