A Completed Publication about Localizable Sentences Research

by

William J. G. Overington

Published by the author
on Wednesday 21 March 2018

For some years, since 2009, but there are some earlier published notes that are relevant, I have been researching a concept that I have named localizable sentences.

In 2009 I was considering what types of item could be encoded each as a character in the Universal Character Set.

Letters of the English alphabet, such as I am using writing this publication, are already encoded as are the characters used for many other languages. There are also symbols and there is the subset of symbols that are known as emoji.

I wondered if a whole sentence could be encoded as a single character.

I then realized that if a whole sentence were to be encoded as a single character then that character could become encoded into a message and sent electronically to another person and then the character could upon receipt be localized into text in the language of the person receiving the message.

The localization could be done automatically.

Thus there would be communication through the language barrier - only for those sentences encoded in the system - yet communication through the language barrier.

Certainly there would need to be a list of character code points and their meanings expressed in a particular language already available in each computer. Also each sentence that were encoded would need to be grammatically stand-alone, so that the localizing of each localizable sentence would be independent computationally of the localizing of every other localizable sentence.

So the system would need setting up.

There are various balances.

For example, a set of sentences that are supported needs to be chosen. This need not be done all at once. A starting set could be published and more sentences added each year using a consultative process so as to make the capability of the system follow the needs of end users.

However, the more sentences that are encoded, the larger are the files that need to be prepared by people and then searched by computer systems.

Would each thus encoded localizable sentence need a distinctive glyph different from the distinctive glyph of each other thus encoded localizable sentence?

Initially I suppose that I thought that they would, as each thus encoded sentence would be a character in the Universal Character Set.

So I designed some glyphs and indeed made various versions of a font that included the glyphs, each mapped to a character code point in a Private Use Area of the Universal Character Set. That was intended for research purposes: a permanent widely-used system would need formal encoding within the Universal Character Set.

As time has progressed, my ideas for the encoding for the sentences has changed. I no longer consider using individual code points from the Universal Character Set, one of them for each localizable sentence, though I opine that considering that was probably a necessary stage for my reaching the present system.

Later I considered there being one character encoded within the Universal Character Set as a base character and then a tag character sequence being used after that base character.

My present idea is to use a sequence of existing characters from the Universal Character Set, using the integral sign followed by a number of circled digit characters.

In 2016, wishing to popularise the idea of localizable sentences and to consider ways that they could be applied, I decided to try to write a novel based around localizable sentences and their applications.

I had been interested in creative writing for many years. I attended an evening class on creative writing in 1997. After that I wrote a story and published it on the web. This was in web pages and was before I had the facilities to produce a Portable Document Format publication.

So I started the novel, using some of the characters (characters here in the sense of characters in a story, not characters in the sense of the Universal Character Set) from the earlier story.

There was no overall plan. I just got started and wrote, publishing each chapter on the web after it was finished. Sometime later it was suggested to me that an ending was needed, so I wrote and published an ending, though an ending that implied success but without going into it, so as not to prejudge where the research project or the novel might go between where I was in the novel and the ending. The ending was listed as Chapter 80. At the time it was published, in April 2017, the novel was written and published to Chapter 27.

So the novel has progressed, and as of today there are published Chapters 1 .. 27, 29 .. 32, 34 .. 41, 60, 80.

Recently I have become aware that the novel cannot be catalogued until it is complete. Completion appears to be a necessary condition for cataloguing, but is possibly not a sufficient condition for a novel published in chapters with each chapter in a separate document on the web.

So I am in the situation that the nature of the project is that both the novel and the research activity are on-going and thus completion is not regarded as imminent, or even planned, as the idea is that it is an on-going project, continuing and going where it goes.

The webspace has been harvested from time to time by the British Library. However my understanding is that the items in the web pages are not catalogued at the present time. It is not clear as to whether individual items in the harvested web pages will become catalogued at a later date.

As many of the publications are research documents of a few pages and the novel is a continuing project-in-progress and thus not completed, and thus not in the catalogue, an interested reader may not be able to find them easily and also readers who might become interested if they knew of the research might never know of its existence.

This publication is therefore produced as a completed publication so that it will hopefully become catalogued so that information of the existence of the research will become available listed in the catalogue of the British Library.

After this page there are three transcripts of already published documents. I have chosen them so as to give an introductory idea of both the research and of the novel.

There is a chapter of the novel that displays some glyph designs produced some years ago. There is a research document about the present encoding system. There is a chapter of the novel where part of that encoding system is discussed.

So is the novel science fiction? Well, the system described could become implemented with technology available today. Yet the novel does contain some clearly fictional parts which, if implemented in a movie of the novel would require animation sequences and sequences combining animation with real action.

Will the novel influence what happens in the real world?

Thus a real research project and a novel are both projects-in-progress. Yet this publication is a completed publication to point to them from within the British Library catalogue.

* * *

Here is some text from the web page of the novel, added here so as to provide a signpost to readers of this completed publication as to where to find the on-going research and the novel. The webspace has been harvested from time to time by the British Library.

* * *

Chapters of the novel and the author notes have been legally-deposited with The British Library.

The British Library harvests this webspace from time to time, so it is possible that some later chapters could be missing from the latest harvesting that is available at the British Library as they were added to this webspace after that harvesting.

However, the practice is that the author has sent each chapter upon publication as an email attachment to the British Library's Digital Processing Team and received an email receipt and so enquiry to that team as to what is in their archive could possibly lead to access to later documents.

The address of this web page is

http://www.users.globalnet.co.uk/~ngo/locse_novel.htm

and readers may like to check that address first while this webspace remains available upon the web.

The next five pages are a transcript of Chapter 5 of the novel.

This chapter has been selected because it has images of some of the glyphs.

John and Henry are in the computing laboratory and are working out how they will carry out their performance at the Poetry through the Language Barrier evening. They need to get it worked out ready for the rehearsal at La Flava Floro.

For the rehearsal John and Henry are going to play the parts of the visiting performers. At the Poetry through the Language Barrier evening the visiting performers will each speak in his or her own language, but John and Henry will use English at the rehearsals, yet that is not a problem as it is the organization of it all that the rehearsal is about.

"This is tricky." says John "In principle it is all very straightforward, yet when it comes to actually doing it as a performance .... we need to figure this out."

Edith enters.

"Good morning gentlemen."

"Good morning." both reply.

"John, can you tell me about the way that the poetry is to be presented please." says Edith.

"Yes, certainly. I have printed some sheets of cardboard, each showing one of the symbols of the collection of twenty-six localizable sentences that we are using for this event. Twenty-six symbols might sound a lot, but the sentences are in three groups, namely sentences about seasons, weather and colours." says John. "I have done a number of each of them, and in addition I have done two sets for each of the three poems and arranged them in sequence. Henry will stand to my left, and somewhat nearer the audience than me, and he will pick up the cards one by one from the table and hold them out to his right, at about waist height so that the audience can see the one side of the card and I can see the other side. Then I say the sentence localized into English, yet the audience can understand the meaning from looking at the symbol, ... well, if they have learned the meanings from the booklet that they received with their tickets."

"Excellent." says Edith.

"Shall we try one of the poems now please?" asks Edith. "I can be the audience."

"Well, we can try, but there is a bit of a problem at the moment."

"Oh."

"Well, in principle it is all very straightforward, yet when it comes to actually doing it as a performance there is a problem. Henry needs to be able to be displaying the cards both to me and to the audience in a continuous manner, yet having the cards in a pile does not work as sometimes there is a delay in trying to get the next card off the top of the pile. We then tried setting the cards out with about forty millimetres of each card overhanging the edge of the table, but picking up one card can move the position of the next card: a longer table could be used so that the cards did not overlap, but then Henry would have to move along the table as the poem progressed and that would disrupt the performance. There is also the problem that Henry needs to put each card down after I have read it out. We need to figure this out."

“Ah, well, can you try it anyway as best you can so that I can get an idea of how it is intended to be please?”

“Yes, certainly.”

John and Henry take their places, Henry behind the table to the left and in front of John.

Henry holds up the first card in his right hand.

“It is winter.” says John.

Henry places the card down then picks up the next card from the pile and hold it up.”

“The colour is white.”

Henry places the card down and then tries to pick up the next card but has a problem getting it from the top of the pile.

“Oh,” says Edith “we need a better method of picking up the cards.”

“Yes.” says Henry. “It is a bit of a problem.”

“Have you got the poem as localized into English?” asks Edith.

“Yes.” says John as he passes a sheet of paper to Edith.

Edith reads, to herself, not out loud.

It is winter.
The colour is white.
It is spring.
The colour is green.
The colour is yellow.
It is summer.
The colour is green.
The colour is yellow.
The colour is red.
It is autumn.
The colour is yellow.
The colour is brown.
It is winter.
It is cloudy.
The colour is grey.
It is snowing.
The colour is white

“Ah, it goes a full year, through a cycle of the seasons. The various colours ... they refer to different things, even though it does not always say what is the thing to which it refers.” says Edith, out loud.

“Henry, the symbols on the cards that you held up. Can you explain them to me please, are they just abstract designs or is there a clue to the meaning in them, some sort of structure?” Edith asks.

“Well, the first one that I displayed is this one.”

Henry holds up a card.

“That means,

It is winter.”

“It is abstract, yet there is a clue to the meaning. The large shape is common to all of the seasons and indeed the same shape is used for the statements about the weather. The two smaller shapes are representing leaves. For the sentence It is spring. there is one vertical leaf, for the sentence It is summer. there are two vertical leaves, for the sentence It is autumn. there is one horizontal leaf and for the sentence It is winter. there are two horizontal leaves.”

“For the weather there are symbols that give a clue, such as this one for It is snowing.

and for It is cloudy. there is this one.”

“Ah, thank you.” says Edith.

“What about the sentences for colours?”

“Yes, at the moment there are fifteen of them For example, The colour is red.”

“Fifteen!” exclaims Edith. “How do you remember the meanings of all those symbols?”

“Oh it is not too hard. They each have the same distinctive shape at the left, so that shows that the localizable sentence is about a colour. The differences are at the right, and they are based on the Petra Sancta system of shading using lines used to indicate colour in printed diagrams before colour printing was invented, though still used today. However, the Petra Sancta system did not describe every colour in the set of colours, so some are innovative. The sentences each start with The colour is .... so that the word colour, being a noun, means that there is no carry forward of noun gender from any earlier sentence.”

“This is interesting, please go on.” says Edith.

“Well, I suppose if one looks at the designs of the symbols for colours that the colours can be grouped into three groups of symbols - namely those directly from the Petra Sancta line shadings, those that are derived from Petra Sancta line shadings, and the group that goes from black to white through various shades of grey.”

Henry picks up a collection of cards for colours and sets them out on a table in front of Edith.

“As you look at them, the top row are those directly from the Petra Sancta line shadings, red, yellow, green, blue, magenta,

the middle row are those that are derived from Petra Sancta line shadings, namely brown, orange, cyan, pink, sky blue

and the lower row are new designs that go from black to white through various shades of grey, namely black, dark grey, grey, light grey, white.”

“So,” says Edith “Using what you have just told me I could express the poem using just the symbols. I could try to write a poem myself. That sounds fun. .... You mentioned twenty-six symbols being used, so, could you put the rest of them out and explain them please?”

“Yes certainly.

Here are the symbols for the four seasons, for spring, summer, autumn and winter,

and here are seven symbols, one each for raining, snowing, sunny, cloudy, windy, hailing and foggy, each of them starting It is ... , for example, It is raining.:

that is the twenty-six being used for the poetry evening.”

“Yet not every localizable sentence in the system has a symbol does it.” adds Edith.

“That is correct. Early on in the research each sentence being used had its own symbol, but now it is only some of the sentences in the system that have a symbol, mainly those from long ago - they have been done, so why waste them - and a few others that might be used in signs or on cards, yet symbols are not needed for automated electronic systems. Though having symbols is sometimes fun, like being able to have the poetry evening, though of course we need to solve the problem of how to present the cards.” adds Henry.

“I have printouts of the English localizations of the other two poems for the poetry evening if you would like them.”

“Yes please.” says Edith enthusiastically.

Here is the second poem.

It is summer.
The colour is green.
The colour is green.
It is autumn.
The colour is orange.
The colour is yellow.
It is autumn.
It is windy.
The colour is orange.
The colour is yellow.
The colour is yellow.
The colour is orange.
It is winter.
It is foggy.

Here is the third poem.

It is summer.
It is sunny.
The colour is sky blue.
It is summer.
It is raining.
It is summer.
It is sunny.
The colour is red.
The colour is orange.
The colour is yellow.
The colour is green.
The colour is blue.
The colour is magenta.
It is summer.
It is sunny.
The colour is sky blue.

“Thank you.” says Edith.

Edith returns to her office.

John and Henry try to think how to solve the presentation problem.

The next four pages are a transcript of a research document.

The document explains the encoding being used at present in the research project.

An encoding space designed for application in encoding localizable sentences for communication through the language barrier

William J. G. Overington

Thursday 10 August 2017

This space, which I choose to name star space, is entered by a character sequence consisting of an integral sign character followed by three circled digit characters. The space is thereafter exited in certain circumstances.

1. If the first and third circled digit characters are the same yet not a circled zero, then exit is after as many additional circled digit characters as indicated by the third circled digit character.

2. If the third circled digit character is a circled zero, then, regardless of what is the first circled digit character, the next three circled digit characters after the third circled digit character taken together determine how many more circled digit characters there are after those next three circled digit characters.

3. Otherwise, exit is after the third circled digit.

4. In expressing star space within an information technology application, if any character other than a circled digit, a space, a comma, a carriage return, a line feed is encountered before when exit would be expected then exit takes place as an error exception in the expressing of the coding in the particular situation.

Thus many sequences in star space consist of an integral sign and three and only three circled digit characters. Certainly, the way that star space is designed could result in some valid sequences of over one thousand circled digit characters, for example, if a sequence were to start with, for example, ∫②⑤⓪⑨⑨⑧ as nine hundred and ninety-eight more circled digit characters would then be needed to complete the sequence. There is no present intention to use such long sequences and at present there is no known application where that would be desirable. The format is simply chosen so as to minimize the possibility of future developments running out of encoding space.

* * *

The star space is designed for application in encoding localizable sentences for communication through the language barrier.

In application, not all of the sequences available within the space are used.

In particular, any star space sequence that has two consecutive characters the same as each other is not used for encoding a localizable sentence. This is based upon the experience of the present author who at one time had a transaction card with a repeated digit in the number and found that there was often a problem in communicating the

number verbally over a telephone link due to there being confusion over whether one digit was being repeated or whether there were two digits the same one after the other.

Level 1.

In level 1 application, a circled zero is not used. This is because in some circumstances the sequence may need to be spoken using just digits, for example over a telephone link. The words for digits are often taught when a foreign language is taught, yet often zero is not taught. So a person knowing only nine digits in a language could potentially use this system over a telephone link so as to communicate through the language barrier to some extent. Also communication could be by using plastic numbers on a desk, and sets of plastic numbers do not always include a zero.

Also, in level 1, no digit is used twice in any one sequence. Thus there are five hundred and four (equals nine times eight times seven) sequences used in level 1. These are intended mostly for application in communication within medical situations, including everyday communication between a patient and a member of staff in a hospital or a care home. For example, "Would you like a drink of water?" and "Would you like to go to the day room?". For specific meetings a human translator is often arranged: this localizable sentence system is intended for everyday communication and for communication in emergencies. However, some of these five hundred and four sequences are for more general communication, such as "Good day." and "Best regards,".

The encoding does include some extremely long sequences. There is no present intention to use very long sequences, yet these are designed into the encoding space so that if people write software to detect star space sequences in text streams in information technology applications, then if some future development needs a longer sequence then the software will accept that sequence as a valid sequence for analysis.

Please consider the following example of a star space sequence together with its meaning localized into English.

∫⑦⑧① Would you like a drink of water?

As GS1-128 bar code technology is being introduced into National Health Service hospitals in the United Kingdom, a specific format is included in this present document for expressing a star space sequence within a GS1-128 barcode.

GS1-128 barcodes can contain characters, though only from a limited set of characters.

For this research, and maybe for long term practical application, the following format is designated for use of star space sequences within GS1-128 barcodes.

Instead of the integral sign use a sequence of two colon characters.

Instead of circled digits use ordinary digits.

At the end add a colon and a semicolon.

Use GS1-128 Application Identifier 97. That is effectively a Private Use code for use within an organization.

Thus ∫⑦⑧① in star space is, expressed in the way that GS1-128 encoding is expressed in printed text that accompanies the barcode, suggested to be as follows.

(97)::781:;

The internal encoding method of the barcode is not quite like that, but that is the format of the text that accompanies the barcode.

* * *

Beyond Level 1.

An example of the application of sequences that use more than three circled digit characters is to use sequences that start with ∫③①③ so as to encode a collection of localizable sentences relating to enquiring through the language barrier about relatives and friends after a disaster using email, possibly, and desirably, with automated localization in the computer of the recipient of the message. Quite detailed sentences could be used: for example, “The enquirer is the brother of the first person that was named.”.

Also, longer sequences could be used, specified using a sequence within a very sparsely encoded part of star space, so as to ensure that some messages can only be conveyed by a very explicit encoding and could not be received in error by just a one bit error, or a several bit error, during the transmission of the star space sequence for a localizable sentence.

* * *

Here are sequences and meanings localized into English of the sequences thus far assigned in star space in the research project. These sentences are intended as examples so as to provide a catalyst for thought of what could possibly be achieved with a grater number of encoded localizable sentences.

∫①②③ Good day.

∫②⑤① Have you vomited?

∫②⑤③ I have vomited.

∫②⑤④ I have not vomited.

∫③⑧① Do you have pain?

∫③⑧② I have pain.

∫③⑧④ I do not have pain.

∫③⑧⑤ I have pain always.

∫③⑧⑥ I have pain sometimes.

∫③⑧⑦ I have pain, but it is not very much.

∫③⑧⑨

∫③⑨①Where do you have pain?

∫⑤⑦③ Shall we use system ‘five seven three’?

∫⑦⑧① Would you like a drink of water?

∫⑦⑧② I would like a drink of water.

∫⑦⑧③ I would not like a drink of water.

∫⑧①② Would you like to go to the day room?

∫⑧①③ I would like to go to the day room.

∫⑧①④ I would not like to go to the day room.

∫⑧①⑤ Would you like to go to your bed?

∫⑧①⑥ I would like to go to my bed.

∫⑧①⑦ I would not like to go to my bed.

∫⑨⑧⑦ Best regards,

The next five pages are a transcript of Chapter 19 of the novel.

In this chapter of the novel the encoding being used at present in the research project is discussed between two of the story characters in the novel.

Thus the previous document and this document are about the same topic, one formal and one in a fictional scenario.

Please note that whereas the two previous transcripts were each inserted into this document using the original source files of the text that was used to produce the PDF (Portable Document Format) document, the original source file of the PDF document for Chapter 19 was not available and so the following transcript was constructed from the text in the PDF document itself. The font used for the integral sign and for the circled digits is a different font than was used when producing the original PDF document as the original font is not available.

“Hello Edith,” says John “I have been thinking about project 573 and setting down some ideas.”

“Good. What have you got so far?”

“Well,” says John, perhaps a little hesitantly, “I have got this.”

John shows Edith a document with the following text.

∫②⑤① Have you vomited?

∫②⑤③ I have vomited.

∫②⑤④ I have not vomited.

∫③⑧① Do you have pain?

∫③⑧② I have pain.

∫③⑧④ I do not have pain.

∫③⑧⑤ I have pain always.

∫③⑧⑥ I have pain sometimes.

∫③⑧⑦ I have pain, but it is not very much.

∫③⑧⑨

∫③⑨① Where do you have pain?

∫⑤⑦③ Shall we use system ‘five seven three’?

“I notice that you have used circled digits,” comments Edith “is that because there is less chance of confusion with numbers used as numbers as such, nothing to do with this system?”

“Yes.”

“Good idea. What is that symbol at the start of each line? It looks like an integral sign from calculus.”

“Well, actually ....”

“John ....” says Edith somewhat disapprovingly.

“Well, I needed a base character and there is not one specially for localizable sentences at the moment, so I used a character that is available, chosen so as to try to minimize any interaction with other uses of the character. .... I thought that it is a sort of summation, using the numbers together to indicate a particular localizable sentence, ... or maybe I just needed a symbol that I thought would not lead to clashes.”

Edith looks at John.

“There was the alternative of using a specially designed symbol, but the problem then is that it would need to be encoded using a Private Use Area code point. It would look good in documents produced using a font of our own, but if someone else wanted to use the sequence of characters to send a message or even to just produce a list, then unless the person were using our font or one compatible with it then there could be problems, the display just being either of the .notdef glyph of the font that is being used or even worse the display being of a glyph that someone else has mapped to the same code point in another font.” replies John.

“Could that happen in reality, ... the other glyph that someone else has designed for some other purpose?”

“Well, it could. How likely is it, I do not know. I suppose that it might be less likely with some parts of the Private Use Area than for others, but who knows. The thing is, using the integral sign and the circled digits has a capability for a graceful fallback display in a computer system without needing any special font from us. True it depends if that computer system has a font with glyphs for the integral sign and circled digits and if the operating system picks up on that, but, as I say, there is capability for a graceful fallback display.”

“Going back to the integral sign, what are the implications of using a mathematical symbol like that out of its proper .... well, usual, context?”

“Well I thought that the use of an integral sign with circled digits is so different that it is unlikely to cause a clash with any usual usage ....”

“Granted, yes ....”

“Software-wise, well, instead of having a unique base character to specify absolutely that a localizable sentence is being specified, if an integral sign is encountered then the software needs, if I may express the computer function in a human thought manner, the software needs to ask ‘Is this integral sign trying to signal a localizable sentence, or is it just part of ordinary text?’ and test for that by looking at the next character to check if it is a circled digit. It is not as if we would produce software that would automatically treat the integral sign as if it were the start of a sequence for a localizable sentence, the software would

always test the next character, or maybe the next three characters before acting. The test is not absolute, but it may work satisfactorily in practice."

"May?"

"Well, yes, 'may', this is a research project and I am not going to say that it will work satisfactorily in practice. Saying that it will might sound more confident, but I deal in results and evidence and so, at the moment, I say 'may'."

"Indeed, yes, you're right. But why the circled digits rather than the tag digits?"

"Well, the circled digits are all in the base plane, as is the integral sign, and I thought that it might well make things far easier to get started, and avoid unnecessary obstacles that might delay implementation, to use only characters in the base plane."

"Well, you have used that approach and you have made progress, so that is good. Yet please keep the choice of base character under review."

"Yes."

"What about localizable sentences that are not part of project 573, localizable sentences with longer code numbers. Have you any thoughts on the base character for those?"

"Well yes, I thought that I would use the integral sign, the same as for project 573?"

"Will that cause confusion?"

"Well I hope not, it should not do. The thing is, it could perhaps cause more confusion if there are two base characters. If there were two base characters and someone used the wrong base character then there could be confusion in that way. So, well, using the same base character may cause confusion, but, well, hopefully it won't do."

"Yes?"

"Well, I was thinking that it would be a good idea, not essential, but maybe useful in some way as we proceed, that we avoid double digits always, not just in project 573, but that we do allow repeated digits within the code number."

"How would that work?"

"Well, in the example that I showed you for project 573, if you could have a look at that again please, there is no entry for three eight three, so I thought that codes where the first digit and the third digit are the same as each other could be the start of longer codes, maybe have each digit no more than twice in any one code, at least to start."

"Why that restriction?"

"Well, I was thinking. In relation to project 573, whilst eventually it would be good to have localization done automatically using electronic equipment, to get things going, and maybe always in some situations, sometimes the numbers need to be passed manually, perhaps by writing them down. It occurred to me that as, in project 573, as each digit is never used twice in any code, if two people, maybe a nurse and a patient, are using the system then each could have nine plastic counters each with a single digit on them and

always test the next character, or maybe the next three characters before acting. The test is not absolute, but it may work satisfactorily in practice."

"May?"

"Well, yes, 'may', this is a research project and I am not going to say that it will work satisfactorily in practice. Saying that it will might sound more confident, but I deal in results and evidence and so, at the moment, I say 'may'."

"Indeed, yes, you're right. But why the circled digits rather than the tag digits?"

"Well, the circled digits are all in the base plane, as is the integral sign, and I thought that it might well make things far easier to get started, and avoid unnecessary obstacles that might delay implementation, to use only characters in the base plane."

"Well, you have used that approach and you have made progress, so that is good. Yet please keep the choice of base character under review."

"Yes."

"What about localizable sentences that are not part of project 573, localizable sentences with longer code numbers. Have you any thoughts on the base character for those?"

"Well yes, I thought that I would use the integral sign, the same as for project 573?"

"Will that cause confusion?"

"Well I hope not, it should not do. The thing is, it could perhaps cause more confusion if there are two base characters. If there were two base characters and someone used the wrong base character then there could be confusion in that way. So, well, using the same base character may cause confusion, but, well, hopefully it won't do."

"Yes?"

"Well, I was thinking that it would be a good idea, not essential, but maybe useful in some way as we proceed, that we avoid double digits always, not just in project 573, but that we do allow repeated digits within the code number."

"How would that work?"

"Well, in the example that I showed you for project 573, if you could have a look at that again please, there is no entry for three eight three, so I thought that codes where the first digit and the third digit are the same as each other could be the start of longer codes, maybe have each digit no more than twice in any one code, at least to start."

"Why that restriction?"

"Well, I was thinking. In relation to project 573, whilst eventually it would be good to have localization done automatically using electronic equipment, to get things going, and maybe always in some situations, sometimes the numbers need to be passed manually, perhaps by writing them down. It occurred to me that as, in project 573, as each digit is never used twice in any code, if two people, maybe a nurse and a patient, are using the system then each could have nine plastic counters each with a single digit on them and

they could set up a code number on a piece of paper then turn it round so as to present it to the other person in the conversation, then reuse the plastic counters when needed."

"Yes."

"So, extending that to a conversation where a larger selection of localizable sentences is available from which to choose, if each person had eighteen counters, two for each digit, then if we keep the codes so that there is never more than two uses of any digit in a code number, then those eighteen counters would be enough to signal any localizable sentence in that set of localizable sentences."

"Ah yes, I follow what you mean. Good."

There is a pause.

Edith asks "What about the localizable sentences that have a symbol, how would that work?"

"Ah, if an end user chose to have symbols displayed then that would be done by using an advanced format font. It is an established technique for displaying a glyph for a ligature such as ct, so that is a matter of having a font that has the glyphs for the symbols within it, with the font also having information tables to specify the details of each substitution."

"So does that mean that the symbol for the localizable sentence 'It is winter.' does not need its own code point?"

"That is correct when using this system of specifying using a base character and a sequence of other characters, whether those other characters are circled digits or tag characters or whatever. Certainly in earlier tests and indeed in earlier thoughts about encoding then the glyph for 'It is winter.' did have its own code point, and indeed the glyph for the localizable sentence 'It is winter.' does have its own code point, a Private Use Area code point, in the font that I used to produce the cards that were used in the performances at the Poetry through the Language Barrier events, but that is just for convenience at the moment."

"Ah. .... Good, good, it looks like you are making good progress."