

## CHOOSE YOUR WORDS CAREFULLY

Patrick de Gale  
Sage de Gale Information Systems Ltd.  
Tenbury Wells, Worcs, WR15 8PJ

This paper will look in turn at:  
The place and flow of terminology in the corporation.  
The attenuation that terms are subject to in the course of that flow.  
Some of the reasons that meanings get altered while words progress from user to user.  
What actions may help the speedy and correct dissemination of the intended message.

### SIGNIFICANT WORDS FLOW AROUND A COMPANY

We are not here concerned with those we use in everyday life, when the everyday meaning is obvious to all of us. 'Please wash up the coffee cups' is readily understood. To some 'a cup washer' might be a specialised member of the staff of a restaurant. To others it may mean that domed edged disk much used by carpenters and joiners. Note please not just the physical description of the problem item, but a deliberate introduction of 'context' by saying who used it. But who invented that specialised term?

#### The role of inventors

This age of rapid technological advance spawns, probably daily, entirely new artifacts. We can't talk about

the process of metal removal in which electrolytic action is used to dissolve the workpiece metal

so production engineers speak of 'electrochemical machining' or ECM. Some of these new terms come from the universities, but these days as much comes from the industrial research and development laboratories. All kinds of widgets need some words to speak explicitly of them.

Perhaps the lab has passed to the production people a computer network device which they describe as having the following new features:

A router  
Hardware flow control  
A new point to point protocol

A dictionary may imply production are going to have to make a shaping tool for joiners, obtain some domestic plumbing fittings for the bathroom and will have to set up a racecourse in the country for the local nob's hunters.

Once they have mastered the production engineering involved, and the product is in a manufacturable state, a new group becomes involved.

#### The technical writers

Being somewhat more literary, they will be wrestling with pretty words for:

The token ring  
An asynchronous host  
The multiple server network

The concept that an engagement ring is a token of esteem can be mastered readily, and perhaps the host at the engagement party was not able to synchronise the tipping of the bottle with the placement of the glass, but the rules of a game of tennis where special nets are needed to field the services of several opponents are giving difficulty. However the service manuals and the user guide gets written, and it is all handed over to:

#### The training department

Now this department doesn't actually do any training. Its role is to produce more documents, and probably computer material, with which the distributors can train the actual users. They understand they have to be meticulous people, so the proper meaning of:

Disk de-fragmentation  
Transparent user access  
Mouse signal traffic

have to be precisely transmitted. They do see difficulties in explaining how to apply super glue to broken vinyl records reliably, particularly if the potential users are hard to find, having already taken on board the techniques of H G Wells' invisible man. They have assumed that travelling mice leave small dark messages, now to be called signals, indicating their route, which may lead them to the users.

Now we have working another literary group,

#### The advertising and publicity staff

Actually these are the first we have come across who do not see that what they write has to be right. (Note the homophone.) A product that needs the user to take a fortnight's course at £250 a day before he gets a glimmer of what it is all about, needs to be presented as dead simple. They latch on to

A chat window with paging facility

as they seem to understand it, and so the ignorant public is bound to be able to.

However

Multivendor interoperability over synchronous  
links

sounds so good it has to go in untouched, unexplained.

So the product is ready to launch. Ha Ha! The marketing plan says 'World wide', doesn't it? And it all starts all over again in seventeen European languages, let alone the rest of the world. If translation is separately organised for each country, as does indeed happen, there are another seventeen departments needing to understand what they have been sent.

To recapitulate the course of information, look at it as a flow chart. (Fig 1.) An indication of the result for the user is summed up by Mark Whitehorn writing in 'pc user'

I'm closer to being a network freak than most,  
yet I find some networking documentation  
impenetrable . . .

My interpretation of some of the terms I have quoted may be far fetched, but the fact that there is a problem at all, is genuine. We need then to look at possible reasons for misunderstanding.

### Novelty

Companies try hard to steal a march on competitors by adding a completely new device to the latest model. Occasionally there is invented a completely new machine, for example a video cassette recorder. That has to be called something. One group has reduced that five morpheme term to VCR, another uses 'video', which is certainly easier to say, even if the etymological root is largely lost, and so there is no innate clue to the meaning in the term commonly used. It is this lack of meaning to which I am trying to draw attention. The chemicals business has the same difficulty. Someone invented

2,2-di(p-chlorophenyl)-1,1,1-trichloroethane

which is scarcely a good name for an insecticide. I do not know who called it DDT, but it does not seem to me a very meaningful name either, even if we would rather not mention it at all on ecological grounds. I strongly feel that the use of initial letters is not to be encouraged.

In contrast the recently developed videophone has an accepted name that is readily understood from its root morphemes.

It is obvious that a newly invented word, or a term comprising common words in a new noun phrase, needs explanation, but a word given a new or restricted meaning may well need to be explained. By now most of us can recognise when 'mouse' is not a rodent, but are we sure what 'socket' means in client-server technology.

Devices are not the only concepts that need new terminology, even if we extend 'device' to include actions such as procedures, ('flow control'), collectives like 'style sheets', and similar intangibles. There are some verbs that get re-adopted, even intransitive ones. We saw 'to chat' earlier. Adjectives too get new meanings. 'Transparent' conventionally implies the use of vision, and when it doesn't it means apparent when perhaps the person you are seeing through would rather not be giving himself away. In computing however it is reversed to mean that the user is 'unaware' of intermediate processes, which themselves set out to make themselves seen through. In fact there may not even be a user involved at all, it all happens 'by mirrors' as the wits say.

### Geography

All chosen words by themselves are not the only reason for difficulty. If our development staff had been in California, and the tech writers were in Europe, they would not be meeting over coffee. Electronic mail is certainly helping, but it is sometimes difficult to express not just that one does not understand the words given, but that the construction is unfamiliar. People in different countries, even if they nominally speak the same language, can misunderstand each other about everyday needs, let alone a novel technology. Let us consider the word 'through'. An American telephone operator asks if you are through if she thinks you may be finished, but her English counterpart is only asking if he has managed to connect you.

The Oxford English Dictionary offers two main definitions of 'through': The first, 'from end to end', makes the Americanism 'from Monday through Thursday' perfectly valid, but many English, expecting to hear 'from Monday to Thursday', follow the example quoted in the OED, 'through the gate'. To them, 'out the other side' is implied and so they would be happy to go through Thursday to Friday.

Translators are not usually native speakers of the source language, so no one can complain if they cannot discern differences of this sort.

### Personnel

The timespan of development and life of a product is enormous. Ten years is typically the time to design in the aerospace industry, and then the sales go on for ten more. By the standards of the UK computer industry, each post in the development team will have had an average of four incumbents. Continuity goes. The original documents are now being written by new people with new backgrounds and literary style. The readers, having changed, no longer have the accumulated experience that came from asking questions, and indeed from having got it wrong here and there and been rudely acquainted of that sad fact. The problems of novelty are going to get repeated.

Documentation systems

Problems are not confined to people. Documents are expected to be transferred and read by machines. Some people are none too bright. Machines are all dead stupid. What might have been sorted out by an astute reader will probably never be noticed by a machine.

One would hope that a corporation settles on a word processing or electronic publishing system to be used worldwide. Nobody here needs telling that if they don't there are going to be some delays and losses in conversion. However here are a few good reasons for that not to happen, at least in the short run:

The company may have expanded by takeover, and the new offshoot was using something else.

The system previously used was really rather old fashioned, and perhaps competition is presenting much smarter material which has to be bettered.

The present system just cannot cope with the extended, or indeed completely different, character set for the languages of the countries now being targeted.

The material may have to go to an outside agency.

Given the real probability that different EP systems are going to be seen, what is likely to happen? Ignoring possible losses of less used characters, like binding spaces, whose internal representation differs, we may see some linguistics related problems:

Effects of typesetting tags. The original material may have embedded in it some characters to change font. The corpus analysis and computer aided translation systems may have been set up to hide the non-language parts of this:

```
<P24M%-2>V<%2>ALUE<%0> is 100<F128>W<F255P255D>
```

which is to display as

**VALUE is 100Ω**

The V and the A are kerned together, and the tags would prevent a dictionary lookup system from finding 'value'. There is an example of the font change problem too.

Index entries. An author may have embedded an index entry inside a term, perhaps to ensure it returns the right page number from a long sentence.

```
When using the point to <$Iprotocol;point to
point>point protocol . . .
```

'Point to point' in the text is interrupted, and may not be seen by the available dictionary lookup system.

Termbank access If the author's publishing system is to be used, it may not be able to access the dictionary or terminology

system in which all the information and guidance that has been built up is locked. This may be because the hardware and operating systems are different; Unix on Sun perhaps instead of MSDOS. Even within MSDOS Windows may be required for the termbank, and the word processor does not support object linking and embedding.

These examples refer to looking up in databases, but of course it is just as important to be able to add to them. This will be an equivalent problem.

### Problems and Feedback

The difficulties we have looked at have to get resolved. This is done by asking questions. In a very small company, working only in one language, this happens face to face or over the house telephone. The asker gets a positive solution. A company does not have to be enormous for this to be impracticable. Let us re-look at the information flow chart, fig 1. Fairly simple.

Now add the lines of communication that could be set up for query resolution, fig 2. Even if the asker could identify the right person at the end of each problem link, the latter does not want to be at the beck and call of all the translators in all the languages, as well as the various documentation teams that have written the original material. Anyhow he may be in bed when the call comes in. Some form of central fielder of questions is essential. The flow chart can then be somewhat simplified to fig 3. There are probably two levels of information provider needed, one near the developers, and the other to fan out the news to translators, who will inevitably be asking the same question many times.

### Management role

It is management's responsibility to deliver the product at the best cost, on time, and of the highest quality. All the difficulties we have been discussing will impact those objectives. Take them in turn.

Cost Time is money. Engineers being diverted from designing to answer questions are costing money. Translators having to ask questions are costing themselves money if they are on fixed price work, or the client money if paid by the hour.

The infrastructure to support mail systems may be lost as an overhead, but the marginal cost of heavy query traffic is still money. Programming, and perhaps a query database, to lay on an easy-to-use query system for translators to use costs money.

Time Delay can cost even more money, and possibly the raison d'être of the corporation. If the product is made available to the sales force late, they lose sales to competitors. Those sales are never recovered. If the product is much too late, the company may never recover its market share, and falter. The money we are talking here is much more than the internal costs of query support.

Quality. The product has got to be right. The documentation of a product, (and of course the words in the user interfaces of systems products), are a part of the product. It is not just the boxes. There is foreign language documentation in the field for well known products which is actually wrong. I do not mean that it has not been expressed as well as it might, I mean the instructions are wrong. Part of the reason for this is that the unfortunate translators do not understand what they are doing. A poor product does not get repeat sales.

The kinds of failures being described are not easy to cost. All the activities tend to get lumped together in the primary activities of the participants. A translator does not make a timesheet entry every time he or she buzzes off a mailnote. But management does need to have an indication of what is happening. If they insist on seeing, and evaluating, reports from a query control system, they may see the scale of their own difficulties.

#### Query control

Queries must be counted. If all questions are routed through a central response point, this is simple. If the questions always come on an electronic mail system, analysis can be automatic.

Some questions will have nothing to do with the words in the material. There will be problems with the publishing system, or data lost in transfer, and so forth. The bulk will however be questions of understanding. A few of those will be a need for help on the grammatical structure of the original, and similar problems with style. The rest, and easily the main part, will be a need to understand the terminology. This has to be brought under control.

#### Terminology control

If problems are ascribed to woolly understanding of the terms in the company documents, those who generate them are going to have to explain them, and then stick to those terms in the same sense and not use synonyms. There are less and more extravagant ways to do this. The right one is really only to be identified after some analysis of the problem. They include

Glossaries Simple lists of the new terms in a document ought to be any author's habit. In a larger organisation different people ought not to define different meanings for the same words. Sometimes this is inevitable, as many words have accepted meanings in the context of a technology, and a different one in a related one. A central company glossary would not only reduce the chance of redefining a term, but also show the environment it was created for.

Term banks The printing, distribution and maintenance of glossaries is itself a significant cost. From the quality viewpoint the maintenance is the most significant. If a developer cannot broadcast he has just invented a super multi-synchronous widget someone else will, and their readers will not

have any idea what they are talking about. A dynamic corporate termbank seems to be the only solution.

Of course we have already mentioned the interfacing difficulties this may entail, but forewarned is forearmed.

Translators - Urgent

Work hard on your managements to get something in place.  
In 1994.







