

LinguaNet:...SETTING A LINGUIST TO CATCH A THIEF

Edward Johnson

Wolfson College and The Judge Institute-University of Cambridge

Frontier controls in Europe are fading, some have vanished. Police communication across those frontiers is not however keeping pace with the changes. Whereas criminals and their spoils now move with greater freedom, the transmission of cross frontier police messages concerning such movements are hindered by a multitude of linguistic, procedural, administrative and legal impediments. The European Union is policed by some 1.4 million police officers. Within the fifteen countries, over 100 independent police forces operate in 18 languages. Many messages which might help put criminals behind bars, restore property to its rightful owners and reunite family members are simply not attempted. Although there are organisations such as Interpol which can aid police officers, they are not generally set up to handle queries "in real time". This paper introduces Test-Bed LinguaNet a project in multilingual, multimedial cross border police communications. A prototype special purpose network has been installed between the police of several European countries. The purpose of the project is to extend this prototype with particular respect to the linguistic facilities and the connectivity which can be offered to the users. A similar paper was presented at CLAW '96, KULeuven Centre for Computational Linguistics, in March of this year.

INTRODUCTION

The purpose of this paper

The project in question, Test Bed LinguaNet, is problem centred. It is prompted by the general need for better cross border communications between the police, emergency services and related organisations of Europe. It is driven at close quarters by the specific day to day requirements of European police units and its target is the production of a multilingual operational system which can be put straight to work. To achieve this a wide range of disciplines are invoked simultaneously. The content of this paper is not therefore solely linguistic and whilst controlled language methodology plays a large part in the project I will not attempt here to argue the merits of controlled language nor to discuss the pros and cons of individual controlled language schemes. The objective is simply to introduce, as a whole, a communications system-build with a practical objective in the hope that its approach, structure and orientation may interest researchers in appropriate fields. Equally, the project and the system it will produce may provide an application platform for facilities being produced by other researchers. It may even encourage such endeavours. In this regard the paper may be read as an invitation.

The technologies which concern the project's participants are those enabling the controlled composition, manipulation and consequent exchange of text elements; the production and transmission of voice output from text; system responses to voice input; the preparation and transmission of graphics; the maximum practical security of communication and the integration of these separate features into a working whole. Carrier technologies include radio. The key linguistic methodologies are controlled language message composition and translation via user specified templates; corpus derived multilingual lexica and discourse

protocols for operational speech and text languages together with restricted vocabularies for discrete communicative functions.

The Programme and the Consortium

The current phase of LinguaNet development is supported by the European Commission's Framework IV Programme - Telematics Applications - Language Engineering under a project entitled 'Test-Bed LinguaNet'. Over the next two and a half years, three 'industrial' partners: ProLingua, Philips and Kent Constabulary will work with four academic research groups at Cambridge (Judge Institute), Leuven (Faculteit Rechtsgeleerdheid), Copenhagen (Handelshøjskolen) and Bordeaux II (Département de Langues Vivantes Pratiques)¹. A user group of 15 European police forces in five countries which is linked by way of a prototype version of the LinguaNet system is coordinated by Kent County Constabulary. The project is funded on a cost sharing basis, the EC contribution being approximately 1.5 million ECU.

Resources

The main resources of the project are of course the technical, linguistic, legal, managerial and professional skills of the partners' researchers. There are however some particularly pertinent resources which should be mentioned. The listing and explanation of these form a simple structure for the rest of this paper.

Operational Communications Research

To begin with the project has the benefit of a considerable body of relevant past work. This includes domain and corpus data, know-how, experimentation and feedback resulting from fifteen years of research and development work in operational communications. This began in 1981.² It comprises the design and implementation of controlled languages, communications protocols, and purpose specific lexicons for SeaSpeak, Vessel Traffic Control and PoliceSpeak, the study of bi-national Inter-Agency Communication (fire-rescue-police-ambulance-medical etc.) for the Channel Tunnel and work done in preparation for the pilot training programme Airspeak. Several of the researchers who worked on these projects also designed, in the mid 'eighties, the controlled language grammars for the English/French machine translation system 'Linitext' which was built in Cambridge in collaboration with British Telecom. The operational communication projects generally required the application of linguistic controls such as reduction reinforcement and standardisation to the language used in specific operations in an attempt to reduce ambiguity. The resultant languages involve lexical, syntactic, and discursal controls and are medium sensitive, domain specific, and topic centred. Their specification extends however well beyond the conventional boundaries of language engineering into the design of physical settings and implementation of technologies contrived to reduce to a minimum³ the need for risky linguistic interaction.

¹ The European cross-frontier communications problem is too pressing not to have attracted the attention other research teams. In that regard the LinguaNet team are expected to work in concert with current and related projects such as Aventinus, Emmi and Dream.

² Operational languages are treated here as a sub category of operational communications. The term is used wherever a language circuit is a requirement in the coordination and control mechanisms of a technical undertaking.

³ For examples of this see Johnson E. Proceedings of the International Language Symposium Volume 4 Paris

On the subject of risk, it is important to stress that LinguaNet is not an attempt to apply to operational police communications, the relatively complex grammars, huge lexicons, glossaries and thesauri designed for MT or authoring systems such as Linitext. Of course it is tempting to try to achieve from non speakers of the French language, police messages of the reasonable, if stilted, quality shown by the following business correspondence example of the sort that were produced on Linitext as long ago as 1988 by monolingual (English) secretaries.

..... *Je vous écris en réponse à votre télex du 24 juin à propos de notre dernière commande.*

Nous vous prions de faire suivre le tissu d'ameublement 'anti-static' sans retard en petite vitesse.

En attendant de recevoir la livraison à temps pour notre conférence et en espérant être en contact avec vous assez tôt.

Je vous prie d'agréer.....etc. etc....

No machine translation facility exists which can be trusted in operations which may be a matter of life or death and for that reason LinguaNet is restricting itself to relatively humble linguistic devices.

The LinguaNet Prototype

The second major resource for the work is a dedicated communications network installed by Prolingua Ltd. between a number of operational locations in 1993. This simple prototype (see 1995 map below) is in daily use and now connects twenty police units in five European countries as well as the emergency control centres at either end of the Channel Tunnel. The prototype provides a continuous stream of genuine individual messages and complete dialogues in several languages for analysis.

The first version of the prototype was effectively divided into two layers: the user-interface and translation software, and the electronic mail system. The top layer was written by Prolingua Ltd while the lower layer used standard Unix components. This split was deliberately made to allow for the use of alternative mail transport systems.

The top layer is a single program which contains the user interface and the translation software. It was first written in standard ML and used the Poly/ML system written by David Matthews at the University of Cambridge with the X-Windows/Motif interface developed by Abstract Hardware Ltd. This layer interfaced to the electronic mail system by running Unix commands.

In order to minimise costs and for ease of maintenance, the LinguaNet software programme is run on standard IBM compatible PCs with commercially available modems and printers. To operate the system requires a standard analogue telephone line with the ability to

1989 **Les langues et la concurrence économique** also published in: **Fachsprache** International Journal of LSP Vienna 1-2 1990 **Language and Economic Competition**

send and receive calls from the international telephone network. This can be achieved by use of a dedicated line or via a dedicated extension served by an automatic switchboard.

From the outset, the aim of the development of the system has been to produce a system which is simple to use maintain and install. The software is already configured to run English, French, Dutch/Flemish, German, Spanish and Danish versions. The project team are currently working to provide user interfaces in other European languages in preparation for further extensions of the network.

Credible user needs seldom come from investigations conducted in the absence of any of the available or predicted technology. Without this type of exposure systems designs run the risk of being too closely attached to a 'metaphor' of pre-existing working practices. Worse, a 'solution' is devised which is largely technology driven and which then has to be accommodated by users.

The presence of an early version of the predicted technology permits the involvement of users at a very early stage in the development process. For the purposes of this project we therefore speak in terms of two systems: one, the **preliminary system** for identification of user needs, initial validation of purpose-built modules, and for providing large databases of genuine communications for analysis, and two, the **object system** which will incorporate certain additional tools and facilities which have been identified in the first stage and then translated into functional specifications from user needs and exposed to testing on the prototype by those users. The project maintains the option of adopting the platform provided by the preliminary system as the platform for the object system but this is not a requirement

In the LinguaNet project there are natural controls upon the system build brought about by the urgency of the cross border requirement. The aim is to provide, quickly, something more functional than the prototype and it is unlikely therefore that we can, in this relatively short term, address all the user needs which may be revealed by a user requirements analysis. Equally, it is anticipated that whilst some of the apparent needs will be capable of translation into functional specifications for telematics tools which are technically feasible, such tools do not yet exist in a sufficiently reliable form for implementation in this setting. We may therefore have to forego for the time being some of the more exciting developments in the field and restrain ourselves to selection, adaptation and improvement of existing technologies and methodologies.

The core user group

The third resource of particular note is probably the most important of all and arises in part from the existence of the prototype. It comprises the people using it. The police forces on the network have formed a core user group which meets regularly to discuss problems and review developments. It is also the forum at which ideas for additional facilities for the system are proposed and examined. It grows by the day and its current composition (of forces) is as follows:

Belgium:

Rijkswacht/Gendarmerie

Police Judiciaire/Gerectelijke Politie

France:

Gendarmerie Nationale

Renseignements Généraux: Police Nationale

DDCILEC: Police Nationale - formerly PAF

Denmark:

Interpol Copenhagen

Netherlands:

Korps Landelijke Politiediensten

CRI Den Haag

Rivier Politie Rotterdam-Rijmond

Politie Zeeland

UK:

British Transport Police

Hampshire Constabulary

Suffolk Constabulary

Sussex Police

Devon and Cornwall

Kent County Constabulary

The wider user group

The development, whilst nurtured by the police, has potentially much broader applications. The police application is first and foremost a good developmental environment since a solution is being actively sought by a widely distributed yet disciplined community using different languages but which shares a common purpose. Even so, opportunities for migration to other sectors are evident within this application. For example, the weakening of frontiers means that we can now envisage more collaboration between Europe's emergency services. At present however Europe's fire and medical response units are impeded in their ability to work together by the same communications difficulties which afflict the police. With respect to crime it is the credit card companies, vehicle rental companies, insurance companies, vehicle security organisations and others which are now expressing interest in the development. The initial interest of course arises from the thefts, frauds and consequences of accidents from which these organisations suffer. In that regard we expect to enlist their technical collaboration in experimenting with specific modules which cope with their particular concerns. A natural development from this is the transfer of closely related multilingual systems which serve their commercial communications too.

Wider applications suggested by the above are being explored by way of an 'outer shell' interest group which comprises representatives from the following:

CLOSE AND DIRECT

RELATED AND DIRECT

INTEREST INDIRECT

**OTHER DGs
(DG XIII's OWN USER GROUP)**

OTHER POLICE SERVICES

EMERGENCY SERVICES

EMERGENCY PLANNING

EXISTING INTERNATIONAL

POLICE BODIES

SECURITY COMPANIES

CREDIT CARD COMPANIES

VEHICLE HIRE COMPANIES

INSURANCE COMPANIES

VEHICLE SECURITY COMPANIES

BANKS

DG7 - TRANSPORT

DG8 - DEVELOPMENT

DG11 - ENVIRONMENT NUCLEAR SECURITY AND

CIVILPROTECTION

Police and Emergency Service communications equipment

Additional technical resources are expected to come from Philips, a company which has a long record of supplying communications equipment to the police and other public services. This fourth and particular contribution comes into play not only with respect to technologies which extend the system functionality, such as permitting text messages to be reciprocally glossed and supported by images, but technologies which extend connectivity. Important aspects of front line police work take place not in offices but on the street and it is essential therefore that any communications system built for it has adequate mobility. Already, a simple prototype exists within ProLingua which converts controlled and translated police text messages to speech and one aspect of the project is to integrate technologies to deliver that speech to where it is wanted. Philips mobile radio and paging divisions are active in providing this.

The question of the law

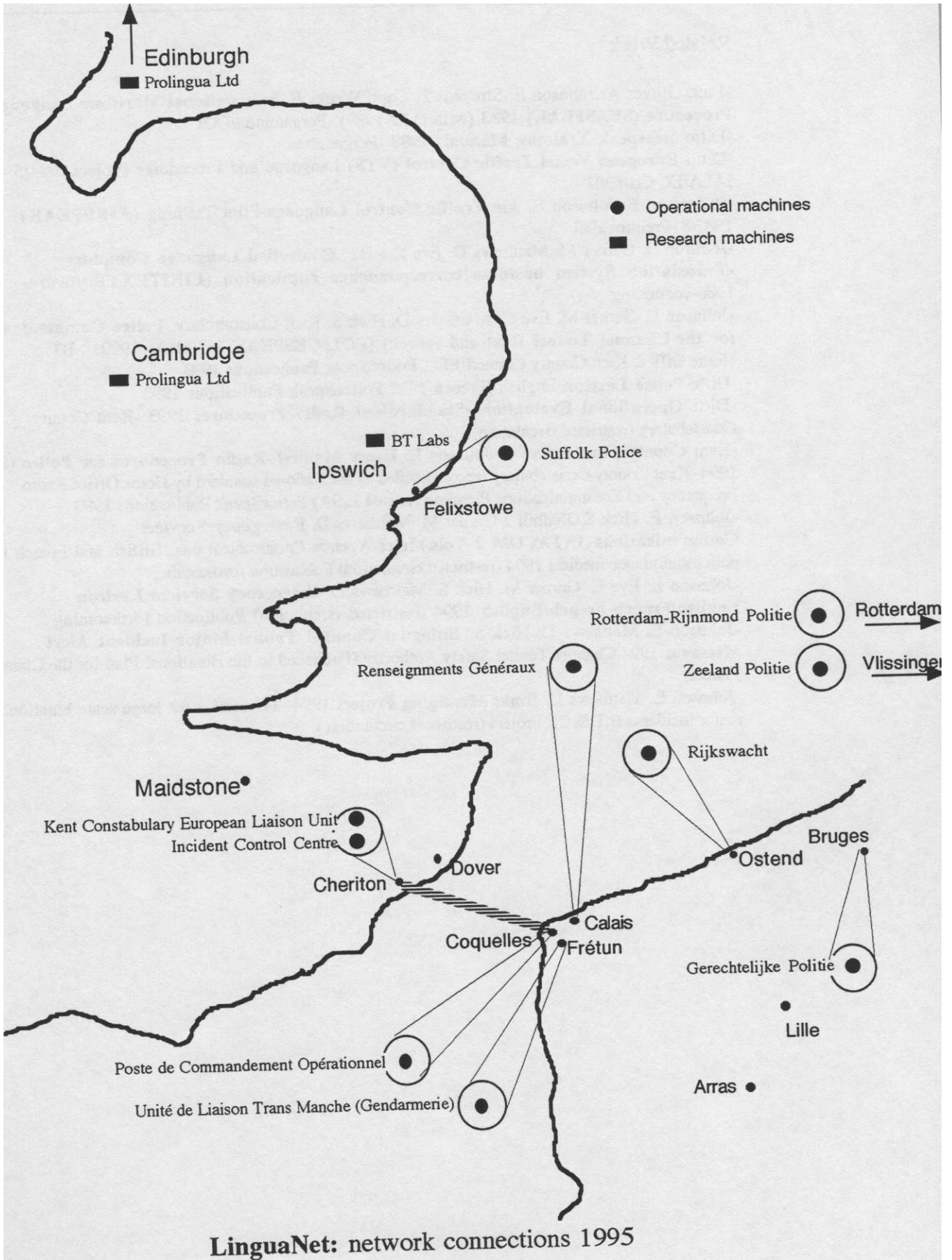
A fifth resource arises from a special obligation to establish the legislative constraints which govern cross border communication. In an operational environment involving the police not all that *could* be communicated *can* be communicated, not all inter-system links which *could* be made *can* be made and requests which are commonplace in one administration are not permissible in another. Such considerations override technical and linguistic feasibility.

This does not mean that the project need limit its research endeavours to what is currently permissible. Moves such as the Schengen accord, the creation of Europol and the building of the European Information System in Strasbourg are clear indicators that European laws pertaining to police communication will change. We need to understand the likely direction of these changes in order to tailor our objectives. Furthermore we must be fully conversant with the present constraints in order that the system which emerges from this project can be activated as soon as it is ready. For a comprehensive study of these issues we are relying on our partners at the criminology department of the University of Leuven where Professor van Ootrive and his team has a distinguished record in the field.

CONCLUSION

This paper, declared as little more than an announcement and an invitation, has briefly outlined a project which is achievable. Its initial methodologies may not appear to be very exciting to linguistic researchers wrestling with exophora but its application is evident and already manifest. The most oft quoted exercises in controlled language applications, weather forecasts, maintenance manuals and the like are evidence that real world, purpose directed projects (in the right hands) represent an increasingly respectable *modus operandi*.

The operational communications dimension may add something further. In these professional settings we have the advantage of access to the entire communication process. We have the speakers and the spoken to in the same time frame. More than that we have instant access to real life interpretations of the messages exchanged. As a result of them planes climb, ships speed up and firemen rush into doorways billowing with smoke. Outcomes such as these greatly increase the validity of judgements of the quality of engineered linguistic systems. Whether linguists are equipped to bear the consequences of linguistic failure in contexts such as these remains to be seen.



Related Work:

- Ltnt. Glover A. Johnson E. Stevens P. Capt. Weeks F. **International Maritime Language Procedure (SEASPEAK)** 1983 (with IMO 1987) -Pergammon/DOI
- Ditto **Seaspeak Training Manual** 1988 -Pergammon
- Ditto **European Vessel Traffic Control (VTS) Language and Procedures Project, 1984/5** - IALA/EC Cost 301
- Robertson F. Johnson E. **Air Traffic Control Language-Pilot Training (AIRSPEAK)** 1987/8-Prentice Hall
- Johnson E. Garner M. Matthews D. Eve C.+ BT. **Controlled Language Computer -Translation System business correspondence application (LINTEXT) software-** 1986-continuing
- Johnson E. Garner M. Eve C. Matthews D. Hick S. Kent Constabulary. **Police Communications for the Channel Tunnel** (text and speech) (**POLICESPEAK 2 Vols.**) 1992 .-BT, Home Office, Kent County Council/EEC. PoliceSpeak Publications 1993
- Ditto **Police Lexicon** English /French 1992 PoliceSpeak Publications 1993
- Ditto **Operational Evaluation -Standardised Radio Procedures** 1993 -Kent County Constabulary (restricted circulation)
- Kent County Constabulary + Johnson E. **Users Manual -Radio Procedures for Police (SRP)** 1994 -Kent County Constabulary (recommended as the National standard by Home Office Radio Frequency and Communications Planning August 1994) PoliceSpeak Publications 1993
- Johnson E. Hick S.Gledhill J Garner M. Matthews D. **Emergency Services Communications (INTACOM 2 Vols.)**-Inter-Agency Communications: British and French fire police ambulance/medical 1994 (restricted circulation) Publication forthcoming
- Johnson E. Eve C. Garner M. Hick S. Matthews D. **Emergency Services Lexicon English/French-French/English** 1994 (restricted circulation) Publication forthcoming
- Johnson E. Matthews D. Hick S. **Bilingual Channel Tunnel Major Incident Alert Messages** 1994 Channel Tunnel Safety Authority (Published in the Binational Plan for the Channel Tunnel)
- Johnson E. Matthews D. **Binat Messaging Project** 1994- Telematics for large scale binational major incidents B.T./KCC Project (restricted circulation)