Building Multimodal Systems: Compromise between Theory and Practice

Marilyn CROSS†, Christian MATTHIESSEN††, Licheng ZENG††, Ichiro KOBAYASHI†††,

†DSTO - C3 Research Centre
Department of Defence
Canberra
marilyn.cross@dsto.defence.gov.au
††School of English and Linguistics
Macquarie University
Sydney
cmatthie@laurel.ocs.mq.edu.au
lzeng@laurel.ocs.mq.edu.au
†††Hosei University
Tokyo
koba@mt.tama.hosei.ac.jp

Abstract

In the context of building a system that assists knowledge workers to process multimodal information sources in the domain of communicable diseases, the compromise between the theoretical ideal of a unifying representation and the compromises required for the pragmatics of the application will be discussed. The HINTS application assists information analysts to retrieve relevant documents from multiple sources, extract information form those documents and generate multimodal presentations in areas of interest. The foundation research question that guided the research and design of the prototype was the possibility of unifying semantics across different semiotic systems which are instantiated in different modalities, viz, the linguistic versus the visual semiotic for this exploration. For the application domain in which language carries the wider range of meanings and where the visual semiotic is complementary, the premise was explored that a theoretical model of the semiotics of language might be used to unify the semantics across different modalities. An analysis of the domain showed the premise was viable. The subsequent design and implementation of the prototype highlighted the dialectic between meaning potential and instantiation and how the change in balance from retrieval, to extraction and to generation needed to be managed computationally, as well as theoretically.

1.0 Introduction

On any given topic there are numerous information sources produced from diverse perspectives. This information may be structured (databases), or unstructured (maps, text, images etc) and many of the current tools for searching and processing information are too generic to meet the needs of professionals who are working within domains in which they are highly knowledgeable. Uniting information processes across multimodal sources has been explored in a limited number of domains, cf. foreign exchange rate changes (Kobayashi and Sugeno, 1994) and healthcare briefings (McKeown, Pan, Shaw, Jordan and Allen, 1997). A multimodal information management tool (called HINTS) has been prototyped that assists professional health workers to retrieve and extract information, and to generate reports in the domain of communicable diseases. HINTS is the instantiation of a theoretical approach to processing multimodal information that attempts to unify the semantics of different modalities. In building a real system for users we have had to find both theoretical and engineering solutions to a wide range of research problems, some of which will be explored in this paper. The next section begins with a description of HINTS to provide a context for problem discussion. Section 3 explores the theoretical problem of unifying semantics across semiotic systems and section 4 describes the analysis of the semantics of a multimodal domain. The issues of designing a multimodal meaning base based on the domain are discussed in section 5, and section 6 tackles the pragmatics of marrying meaning structures

across different information management processes.

2.0 HINTS: Multimodal System

HINTS is a system for managing multimodal information in the domain of communicable diseases, such as cholera and HIV. Sources of information for HINTS are open source and include Web sites, for example, that of the World Health Organisation, mailing lists, such as PROMED, and structured databases, such as the epidemiological database GIDEON.

In HINTS, the workflow is initiated by users creating a production requirement consisting of concepts to be expressed, the intended audience of the final product, the intended modalities of that product and possible sources of information to be used in its creation. The system then searches for and collects relevant information and returns this to the user for editing and refinement. Three processes are available to the user:

- Retrieval, in which those "documents" which contain information relevant to a user's information requirement are identified;
- Extraction, which extracts content and converts that content into well-defined structures;
- Generation, where the content that is held in a collection of related well-defined structures is re-expressed using one or more modes of representation.

Central to the system is a meaning base that represents the semantics of the domain of communicable diseases and the manner in which they may be distributed across different modalities.

The HINTS prototype forms the context for discussing the theoretical and pragmatic problems of multimodal systems.

3.0 Unifying Semantics across Semiotic Systems

Multimodal human-computer systems are composed of instantiations from a number of different semiotic systems. One may also argue that as well as being constructed from different semiotics, the completed multimodal product is an instantiation of a semiotic for multimedia (or potential for meaning making) which is a synthesis that is greater than the contributing parts.

It may be argued, and not without contention, that language is the richest and most structured system for making meaning that humans use. At least three recognised schools of semiotics have applied concepts from linguistic theory and description to other modalities: the Prague school (Jakobson, 1971), the Paris school (Barthes, 1967) and the school of social semiotics (Halliday, 1978; O'Toole, 1994; Kress and van Leeuwen, 1996). The focus in the current discussion is on the visual modality and within that a relatively selective choice of geospatial representation i.e. maps.

The semiotic system of language is diverse. As described in systemic-functional theory (Halliday, 1978), it creates meaning along at least three dimensions:

1. ideational: construes our experience of the world around us and inside us
2. interpersonal: a resource for enacting our social world, i.e. our social roles and relations
3. textual: directs the process of sharing information as text in context by providing speakers with the resources to guide listeners in interpreting the information

Making the assumption that an instantiation of the visual semiotic is a piece of social communication, the three metafunctions may be described in terms of the visual semiotic. Maps have been described as representing analytical processes that relate participants in terms of part-whole structures (Kress and van Leeuwen, 1996).

In HINTS, the linguistic semiotic was applied in the domain analysis, in the meaning base and in the generation process.

4.0 Analysis of a Multimodal Domain

The domain analysis for HINTS describes the ideational semantics of multimodal documents in the subdomain of Weekly Epidemiological Reports (WER's) of the World Health

Organisation. The analysis of the domain (Matthiessen, 1996) is described in terms of the sequences, figures and elements that together constitute the ideational representation of the world.

Based on the analysis of the domain it was possible to build a domain model, a sample of which is given in table 1.

Figure Type	Actor	Process	Goal
Doing			
1	**health workers**	**medical procedure**	**patients**
Example:		*have been traced*	*all known patients who have left hospital*
2	**health workers**	**laboratory research**	**samples; viruses**
Example:		*were provided*	*blood specimens from several of these initial patients*
3	**diseases**	**infection**	**people**
Example:		*were infected*	*many health care workers*
Happening			
4	**public; patient**	**phase of disease**	
Example:	*among whom 101*	*died*	
Example:	*they [= persons with Ebola]*	*are already haemorrhaging*	

Table 1: Semantic Configurations in Domain of Communicable Diseases

The table shows four types of semantic configurations (or figures) with examples. Thus in configuration 1, the process is typically one of some kind of medical procedure such as *have been traced* with the Actor being some kind of health workers and the Goal as some kind of patient, one example of which is *all known patients who have left hospital.* The domain analysis showed that in the linguistic representation the figures covered the gamut of available process types.

Further analysis of the visual modalities in the domain, viz. maps and tables, showed that they used a very limited portion of the semantic domain as instantiated in the linguistic description. Maps, tables and graphs in the WHO reports were used in similar ways. They provided quantified Tokens representing some unit of measurement as Value, with the Tokens placed at the relevant Location: place on the map. For example, location on a map is typically realised linguistically by means of a name, e.g. *Georgia* and the incidence rate of a disease, eg. diphtheria by a number, e.g. 5.4. The caption provides the means for interpreting the data provided on the map - *Reported incidence rate (per 100,000 population) of diphtheria, USSR, 1994.*

Using the domain analysis it was possible to build and experimentally populate a meaning base that provides a network of the semantic concepts for language and other modalities.

5.0 Design of a Multimodal Meaning Base

What is a "meaning base"? It is a resource of a kind that is related to other "information bases" in computer science and AI — data bases and knowledge bases. Like these bases, it is a resource that can be accessed by a variety of processes e.g. document understanding, some inferencing, and the generation of responses to user queries. More specifically, such resources are repositories of information that are organised according to certain general categories, such as relations and entities.

The notion of a "meaning base" makes direct reference to the established notion of a knowledge base or ontology (cf. Noy and Hafner, 1997).; but it is intended as a complementary conception, replacing knowledge with meaning. The fundamental principle is that meaning is constructed in language, so the approach to the meaning base is language-based.

Halliday and Matthiessen (1997, pp.2-3) introduce the concept of a meaning base as follows:

"What is the significance of this switch of metaphor from knowing to meaning? A meaning base differs from a knowledge base in the direction from which it is construed. In modelling the meaning base we are building it 'upwards' from the grammar, instead of working 'downwards' from some interpretation of

experience couched in conceptual terms, and seen as independent of language. We contend that the conception of 'knowledge' as something that exists independently of language, and may then be coded or made manifest in language, is illusory. All knowledge is constituted in semiotic systems, with language as the most central; and all such representations of knowledge are constructed from language in the first place. This suggests that it should be possible to build outwards from the grammar, making the explicit assumption that the (abstract structure of) categories and relations needed for modelling and interpreting any domain of experience will be derivable from those of language. The contention is that there is no ordering of experience other than the ordering given to it by language. We could in fact define experience in linguistic terms: experience is the reality that we construe for ourselves by means of language."

In the context of HINTS, it is important to note that the meaning base covers all the "modalities" or semiotic systems involved, although with language as the principal one. To describe the design of the meaning base, it is necessary to describe the Multex generator of which the meaning base forms a part (Matthiessen, Zeng, Cross, Kobayashi, Teruya and Wu, 1998). Multex is organised globally into four strata: context, semantics, lexicogrammar and expression.

The contextual level models the environment in which Multex operates, which includes the context of generating multimodal presentation but also the context in which Multex interacts with the production application, which is in this case HINTS. The second strata is the semantic level which takes as input the conceptualised representation of the production application state and generates a set of semantic objects that represent the state as a specification of meanings that can be realised in a multimodal presentation. The semantic objects are organised as a semantic network which constitutes the meaning base. The lexicogrammatical level takes the semantic networks as input and realises these networks in structures appropriate to the modality of generation, for example, grammatical systems and lexis for language, and map objects for maps. The fourth level is the expression level that takes the string of objects and organises them into a multimodal layout, using a conventional publication specification, which for HINTS is HTML.

6.0 Marrying Meaning Structures across Information Processes

In theory it is possible to utilise the semantic networks of the meaning base as specifications for retrieval, extraction and generation. Pragmatically in HINTS, the meaning base has been utilised fully for generation, partially for information extraction and not at all for information retrieval.

In the best of possible worlds, it should be possible to utilise one data structure that will facilitate the three processes of retrieval, extraction and generation. Indeed, object-oriented data structures or templates have been used for information extraction in a range of domains. For each extraction task, a generic type of information is specified for each of the slots in the template and specific instances are automatically extracted from the candidate texts to fill the slots (Onyshkevych, Okurowski, Carlson, 1993). Because the template design is object oriented, it is possible to utilise multiple subtemplate types which represent related information to the parent template type as well as relationships to other objects.

The information extractor used in HINTS (Wallis and Chase, 1997) utilises templates which are capable of extracting simple facts, for example, *date, location, disease name and number of cases* using pattern matching, with patterns defined by regular expressions. The combination of these simple facts provides a higher level template which describes *disease outbreak.* It is this higher level template that provides the values for the data structure in the meaning base and hence the Multex generator.

Retrieval in HINTS is currently achieved using key words, for example, of which the generic type is *disease name* and *location.* A time period is also specified so that users can specify occurrences of diseases within a particular time period. It would not be difficult to build a

template from the slots of *disease name, location* and *time period.*

Thus it would seem form the evidence presented thus far that there is more than sufficient commonality for the three processes to 'share' the same template structure. This has not proven easy. There is a dialectic between potential and instantiation that is much more complex than appears at first pass and which is difficult to capture computationally cf. figure 1.

The domain model is located midway between the poles of instantiation (potential and instance); it represents a sub-potential of the overall meaning potential (or seen from the other end of the cline, it represents an accumulation of instantial patterns found in presentations of communicable diseases);

The three different processes of retrieval, extraction and generation draw on different balances of the potential of the semiotic system (what is possible) and instantiation (what is produced). Retrieval focuses on instantiation with less recourse to potential. Matching at the lexical level with little recourse to grammatical combination has proven effective for retrieval (Voorhees, 1994). Extraction of low level facts can draw on instantiation with relatively little recourse to potential. However, as the attempt is made to extract higher level facts, more demand is made of the potential to provide the blueprint or mapping for the instantiation. Finally at the generation stage a blueprint from the full potential of the semiotic system is required for generation.

HINTS is a pragmatic reflection of this problem. Template types may be seen to represent at least a portion of the semiotic potential.

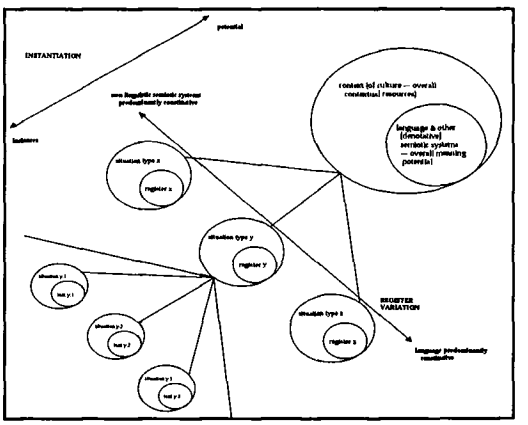

Figure 1: Dialectic between Potential and Instantiation

At the retrieval level no templates are utilised. Keyword search provides the functionality. The possible template that combines *disease name, location* and *time period* is implicit in the design of the graphical user interface (GUI) for the search operation. At the information extraction stage simple templates have been used successfully for low level facts, but have proven more difficult than anticipated for higher level facts. Here automatic extraction has been augmented by user input to the extraction process. For generation, the extraction process has been able to pass to the generator the data structures required which are then interpreted in terms of the potential/ blueprint for generation. The meaning base is interrogated by the generator for the semantic blueprint for generation and essentially provides the semantic potential for generation in an object oriented structure of templates and subtemplates. Thus the dialectic between potential and instantiation is minimal for the retrieval process and maximal for generation. A future goal of the research is to determine how far it is possible to design and access the meaning base to provide support for multiple processes including retrieval, understanding and generation.

7.0 Conclusions and Future Research

In the context of designing and prototyping a system that assists knowledge workers to process multimodal information sources in the domain of communicable diseases, the compromise between the theoretical ideal and the pragmatic solution was discussed. Theoretically and in

terms of the analysis of the domain it was possible to unify semantics across linguistic and visual semiotics, the latter of which was instantiated in terms of annotated maps for the prototype. Future research will explore other modalities that are relevant for the domain, in the first instance, tables and graphs. This in itself requires fundamental work in collating and interpreting multiple instances of facts and subsequently aligning the generation of text and visual representation.

The design and building of the prototype highlighted the tension between meaning potential and instantiation. The dialectic between potential and instantiation is minimal for the retrieval process and maximal for generation. For information extraction there is pattern matching at the instantiation level that is informed by the meaning potential for the extraction of higher level facts. Theoretically it is possible to utilise the meaning base for guidance in retrieval, extraction and generation. Pragmatically, it was possible to use the meaning base substantially for generation and design the templates for extraction following the semantic patterns in the meaning base. The next stage of the research is to explore how better to integrate the meaning base into the system so that it provides responses to the multiple demands of retrieval, extraction and generation. Part of that work will be explored through the vehicle of a language engineer's workbench which will also provide the infrastructure for moving to new domains.

References

Barthes R., 1967. *Elements of Semiology,* London: Cape.

Halliday M. A. K. 1978. *Language as Social Semiotic: the Social Interpretation of Language and Meaning.* London: Edward Arnold.

Halliday M. A. K., and Matthiessen M. I. M. 1997. In process of publication *Construing Experience through Meaning: a Language-Based Approach to Cognition.*

Jakobson R., 1971. *Studies in Verbal Art.* Michigan: University Press.

Kobayashi I. and Sugeno M. 1994. An Approach to Social System Simulation based on Linguistic Information - an Application to the Forecast of Foreign Exchange Rate Changes. *Journal of Japan Society for Fuzzy Theory and Systems, 6(4), pp.10-25.*

Kress G and van Leeuwen T. 1996. *Reading Images: The Grammar of Visual Design.* London: Routledge.

McKeown K., Pan S., Shaw J., Jordan D. and Allen B., 1997. Language Generation for Multimedia Healthcare Briefings. *Proceedings of the Fifth Conference on Applied Natural Language Processing.* Washington: ACL.

Matthiessen M. I. M., 1996 Model of the Domain of Communicable Diseases: Description. Canberra: DSTO Report.

Matthiessen M. I. M., Zeng L., Cross M., Kobayashi I., Teruya K. and Wu C. 1998. The Multex Generator and its Environment: Application and Development. Paper to appear *Proceedings 17th International Conference on Natural Language Processing.* August 10-14, 1998. Montreal.

Noy. N. F. and Hafner C. D. 1997. The State of the Art in Ontology Design. *Artificial Intelligence Magazine, 18(3),* 53-74.

Onyshkevich B., Okurovski M., Carlson L. 1993. Task, Domains, and Languages. *Proceedings of the Fifth Message Understanding Conference (MUC-5),* Baltimore, Maryland, August 25-27, 1993, pp.7-17.

O'Toole M. 1994. *The Language of Displayed Art.* London: Leicester University Press.

Voorhees E. M. 1994. Query Expansion using Lexical-Semantic Relations. *Proceedings of 17th International Conference on Research and Development in Information Retrieval (SIGIR'94),* pp.61-69.

Wallis P. and Chase G. 1997. An Information Extraction System. *Proceedings Australasian Natural Processing Summer Workshop,* Macquarie University, February, 1997.