Table of Contents

Paper 1

An XML architecture for the HCRC Map Task Corpus

AMY ISARD
amy.isard@ed.ac.uk
http://www.ltg.ed.ac.uk/~amyi

Abstract

This paper describes an XML architecture for dialogue annotation, which represents multiple overlapping data streams. Different annotation levels are stored in separate files and linked to a common base level, ensuring that the annotations are maintainable and that changes to one level have minimal effects on another. Some tools and techniques which take advantage of this architecture to allow the annotations to be presented in flexible user-friendly formats are also described.

1 Introduction

Researchers at the Human Communication Research Centre (HCRC) at the University of Edinburgh have a long history of basic dialogue research using vertical analysis on recorded and transcribed data, on which different phenomena are annotated and their relationships determined empirically. For this work, our main resource has been the HCRC Map Task corpus (Anderson et al. (1991)), which consists of 128 task-oriented dialogues with an average length of around 6 minutes. The two speakers face each other across a table, and each has a map in front of them with pictures on it known as landmarks. One speaker, known as the information giver, also has a route marked on the map, and the task is for the information giver to describe this route to the other speaker (the information follower) who tries to reproduce the route on his/her own map. The speakers are prevented from seeing one another's maps by a screen in the middle of the table, and in half the dialogues the screen also prevents the participants from seeing each other's faces.

This corpus has been annotated with a wide range of phenomena including dialogue moves, games and transactions (Carletta et al. (1997)), disfluencies (Lickley and Bard (1998); Branigan et al. (1999)), syntax (McKelvie (1998)), gaze - which records each time that each participant looks up (usually at the other person) or down (usually at the map) - (Boyle et al. (1994)) and landmark references - where a participant refers to a picture of a landmark which appears on the map - (Bard et al. (2000)).

In pursuing this work, we have developed a way of representing multiple streams of data with a common time line and many overlapping hierarchies, while ensuring that the annotations are maintainable, so that changes to one level of annotation have the minimum possible impact on any other. In this paper, we describe our data representation and how we work with it.

2 Design Criteria

Our research methodology has clear advantages in terms of providing a structured framework on which to build, but it raises some serious technical challenges for data representation. The representation must be clear and machine parsable using standard techniques, to minimize programming effort. It must represent

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

amy.isard@ed.ac.uk
http://www.ltg.ed.ac.uk/~amyi.

tree structures because we want to make the hierarchical relationships between elements explicit. For example, a dialogue move, a sentence and a dialogue game may start at the same time, but we want it to be clear that the move is structurally part of the game, whereas the sentence belongs to a different annotation hierarchy. It must allow multiple data streams (e.g. overlapping speakers, gaze, external noises) and if each annotation level is viewed as a tree, it must allow overlapping branches across different levels (e.g. syntax, discourse, and gaze). The representation must also allow for extensions in unforeseen directions, be maintainable, and allow concurrent editing. It must also be possible to display the data in ways tailored to the individual research task, as it is impossible to make sense of all the annotation levels simultaneously.

3 The Base Technology

A number of partial solutions existed in SGML (Goldfarb (1990); W3C (1999)) and XML (Ducharme (1999); W3C (2000a)) for the problems we were trying to solve. The Corpus Encoding Standard (CES) (Ide and Priest-Dorman (2000)) describes guidelines for encoding standards for natural language corpora, based on the Text Encoding Initiative (TEI) (Sperberg-McQueen and Burnard (1994)). The TEI provides methods for allowing overlapping in SGML, but these create a data structure in the form of a chart, and we preferred to work with a data representation based on a set of trees, so we we chose to take a slightly different approach, described in more detail in section 4. Meanwhile, a number of recent developments in the SGML/XML area provided a new approach to creating the functionalities described in section 2. In our structure, each data stream and each level of annotation is stored as a separate XML file, and linking between files is done using a mechanism known as hyperlinking, which allows elements in one file to point to one or more elements in another file. Our implementation of this is very similar to the XLink and XPointer (W3C (2000c)) proposals from the World Wide Web Consortium (W3C), which are close to becoming accepted standards, and when they are established as standards, we will convert our data to conform to them. There are several tools, described in more detail in section 5 which take advantage of this design, including "knit" which makes it possible to expand the hyperlinks within one XML hierarchy to explicitly include the linked elements within a document, or to replace an original element with the links. Stylesheets of various types including XSLT (W3C (2000b)) can be used to transform or display the data. A stylesheet is a document which contains a declarative set of rules for converting one XML document into another XML document, so it can for example be used to create HTML which can then be displayed in a web browser.

4 Structure of the Corpus Annotation

As described above, the corpus annotations are stored in a number of linked XML files. In the Map Task corpus, the lowest level of structure is the timed transcription unit and we have a separate base level transcript for each speaker. Each of these base files contains a sequence of word items, other noises, and silences, which have start and end times which refer to the speech signal. All other annotation is held in separate files, and hyperlinks are used to "point" between files. Timings are generally only included in the base level, but it is a simple matter to "inherit" times upwards through a tree to other annotation levels.

Figure 1.1 shows shows the structure of the annotations for one speaker in the corpus; each box represents a separate file and arrows represent links between files (described in section 4.1). There is a separate parallel set of files for the other speaker. In this example there is another annotation level apart from the base level, gaze, which also points directly to the speech signal. Speakers can look up or down while either or both participants are speaking, so it was not clear that there was a principled way to link gaze to the word files. Therefore gaze elements have start and end time attributes, and when gaze is considered along with other levels of annotation, comparisons are made according to the shared timeline.

The choice of particular annotations is corpus specific - more, fewer or different annotation levels could be used without affecting the general architecture. It is also possible to use this architecture with a speech corpus which does not have timings for each word, by omitting the start and end times from the word-level elements and attaching them instead to a higher level for which timings are available, such as utterances. The architecture could also be applied to a text corpus; in this case, the word level will still be the base

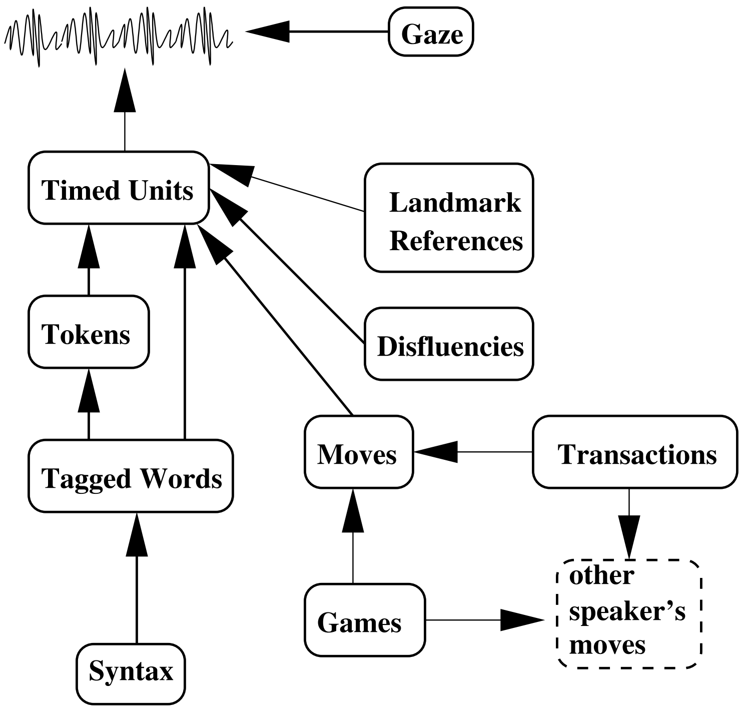

Figure 1.1: Architecture

level and all the other levels will point to it. In this case there would only be one base level file, as there would not be separate speaker streams.

Figure 1.2 shows another view of the corpus, with a stretch of text and several overlapping annotation hierarchies which refer to it.

4.1 Links Between XML Files

In our design, a hyperlink attribute on an element points to a contiguous stretch of elements in a different file. Figure 1.3 shows a) part of a base-level timed unit file, and b) a dialogue moves file with pointers to the base file. In the base file, the elements are of type tu (timed unit) and sil (silence). Each element has three attributes: a unique identifier (ID), start time, and end time, and tu elements additionally have content, which is the transcription of the word. In the move file, each move element has four attributes: an ID, speaker, a label specifying the type of dialogue move, and an href attribute, which provides the hyperlinking. The href attribute consists of two parts, first the name of the file which the element(s) pointed to can be found in, and second a list of ID(s) of the element(s) which are pointed to. If it is a single element, just one ID is listed (e.g. id(q1ec1g.1)) but it is also possible to include a sequence of elements (e.g. id(q1ec1g.4)..id(q1ec1g.14)). In this second case all the elements between the two which are named will also be included.

5 Working with the Data

XML coding makes the structure of a corpus explicit, and this facilitates the process of querying the corpus. There are a number of tools developed by the Edinburgh Language Technology Group (LTG (2000)) which take advantage of the hyperlink semantics used in the corpus. One example is "knit" which, depending on some user-set parameter, expands the hyperlinks either to include the linked elements within the original document, or to replace the original element with the links. This then allows queries to be performed over

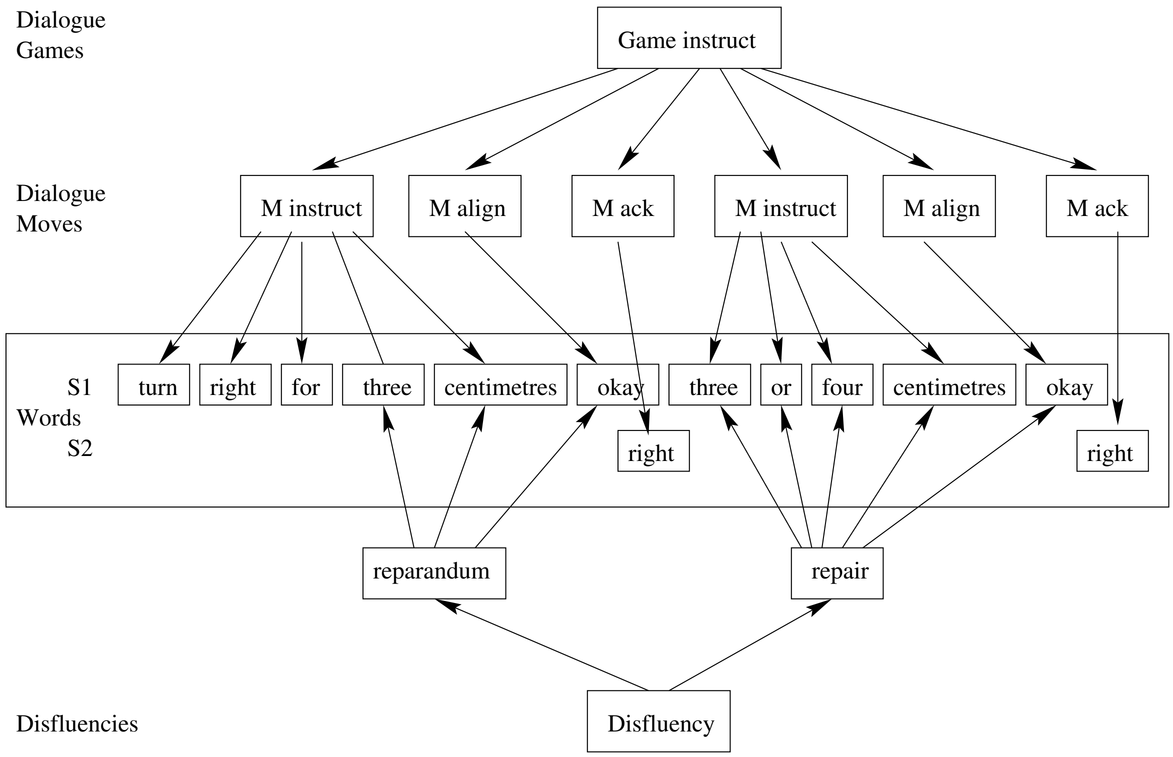

Figure 1.2: Annotation Structure

an entire hierarchy using one of any number or XML query languages (Marchiori (1998)). Knitting and XML querying suffice when a single annotation hierarchy is being processed, but it is more complicated to perform queries across hierarchies. The fact that our each annotation level is linked to the same base level facilitates the formulation of queries which involve more than one hierarchy. For example, one might want to find out whether a particular type of disfluency occurs more frequently during one type of dialogue move than another. The MATE workbench (McKelvie et al. (2001)) takes advantage of the hyperlink semantics, and allows queries and displays to be performed over an entire corpus. It is also possible to perform queries across more than one document using XSLT stylesheets; this is currently rather tortuous but should be simpler once XLink and XPointer become fully standardized.

The XML annotation of the corpus also allows us to provide flexible display options using stylesheets and generic software. The online Map Task Corpus Demo (Isard and Aylett (2001)) shows some flexible display solutions produced in real time using stylesheets to produce HTML, and allows the user to select a number of annotation levels and display formats.

6 Discussion

Our techniques have some disadvantages, in that they require the creation of a large number of files, which are not easily human-readable as they stand, and that tools such as “knit”, MATE or XSLT must be used to work with a whole hierarchy, or a set of hierarchies. However, these are outweighed by the advantages described in this paper.

There are some issues which have not yet been fully addressed; we maintain individual files using standard version control software (RCS) but we have not yet developed a coherent strategy for dealing with the knock-on changes which occur to higher-level files when a base-level file is edited. Multimodality options are also still under development. We can currently link to a speech or video using offset times from the beginning of the signal and start and end attributes on elements. However, there is for instance no software currently available which would allow us to integrate the XML files with, for example, an a)

```
<!DOCTYPE timed_unit_stream SYSTEM "dtd/maptask-timed-units.dtd">
<timed_unit_stream id="tu.q1ec1.g">
<tu id="q1ec1g.1" start="0.0000" end="0.3294">okay</tu>
<tu id="q1ec1g.4" start="0.3294" end="0.8432">starting</tu>
<tu id="q1ec1g.5" start="0.8432" end="1.3702">off</tu>
<sil id="q1ec1g.6" start="1.3702" end="1.5777"/>
<tu id="q1ec1g.7" start="1.5777" end="1.8413">we</tu>
<tu id="q1ec1g.8" start="1.8414" end="2.2201">are</tu>
<sil id="q1ec1g.9" start="2.2201" end="2.3518"/>
<tu id="q1ec1g.10" start="2.3518" end="2.8722">above</tu>
<sil id="q1ec1g.11" start="2.8722" end="2.9644"/>
<tu id="q1ec1g.12" start="2.9644" end="3.0369">a</tu>
<tu id="q1ec1g.13" start="3.0369" end="3.5244">caravan</tu>
<tu id="q1ec1g.14" start="3.5244" end="3.9394">park</tu>
 ...
</timed_unit_stream>
```

b)

```
<!DOCTYPE move_stream SYSTEM  "dtd/maptask-moves.dtd" [
<!ENTITY gfile "q1ec1.g.timed-units.xml">
]>
<move_stream id="move.q1ec1.g">
<move id="q1ec1.g.move.1" who="giver" label="ready"
 href="&gfile;#id(q1ec1g.1)"/>
<move id="q1ec1.g.move.2" who="giver" label="instruct"
 href="&gfile;#id(q1ec1g.4)..id(q1ec1g.14)"/>
...
</move_stream>
```

Figure 1.3: Part of one speaker's base level timed unit transcription and corresponding dialogue move transcription

interface for annotating sections of a map.

No other currently available approach comes close to providing all of the functionality which we require, and because we use standards developed for other purposes, we can make use of many freely available tools, greatly reducing programming effort. We can also make use of new XML techniques as they are developed in the future. It is easy to add new levels of annotation to an existing corpus without reference to existing annotations, and editing of one annotation level has minimal effects on other levels.

Bibliography

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4):351–366.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (2000). Controlling the intelligibility of referring expressions. *Journal of Memory and Language*, 42(1):1–22.

Boyle, E., Anderson, A., and Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, 37(1):1–20.

Branigan, H., Lickley, R., and McKelvie, D. (1999). Non-linguistic influences on rates of disfluency in spontaneous speech. In *Proceedings of the 14th International Conference of Phonetic Sciences*.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., and Anderson, A. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.

Ducharme, B. (1999). *XML: The Annotated Specification*. Prentice Hall.

Goldfarb, C. F. (1990). *The SGML Handbook*. Oxford University Press.

Ide, N. and Priest-Dorman, G. (2000). The Corpus Encoding Standard. http://www.cs.vassar.edu/CES/.

Isard, A. and Aylett, M. (2001). The HCRC Map Task Corpus Demo. http://www.ltg.ed.ac.uk/~amyi/maptask.

Lickley, R. and Bard, E. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech*, 41(2).

LTG (2000). LTXML. http://www.ltg.ed.ac.uk/software/xml.

Marchiori, M. (1998). W3C Query Languages Workshop 98. http://www.w3.org/TandS/QL/QL98.

McKelvie, D. (1998). SDP - Spoken Dialogue Parser. Technical report, HCRC, University of Edinburgh. HCRC/RP-96, May 1998.

McKelvie, D., Isard, A., Mengel, A., Moeller, M. B., Grosse, M., and Klein, M. (2001). The MATE Workbench - an annotation tool for XML coded speech corpora. *Speech Communication*, 33(1-2):97–112. Special Issue: Speech Annotation and Corpus Tools.

Sperberg-McQueen, C. M. and Burnard, L. (1994). TEI Guidelines for Electronic Text Encoding and Interchange (P3). http://etext.lib.virginia.edu/TEI.html.

W3C (1999). Standard Generalized Markup Language (SGML). http://www.w3.org/MarkUp/SGML/.

W3C (2000a). Extensible Markup Language (XML). http://www.w3.org/xml.

W3C (2000b). Extensible Stylesheet Language Transformations (XSLT). http://www.w3.org/TR/xslt.html.

W3C (2000c). XPointer and XLink. http://www.w3.org/XML/Linking.