Querying multimodal annotation: A concordancer for GeM

Martin Thomas
Centre for Translation Studies
University of Leeds
UK, LS2 9JT
m.thomas@leeds.ac.uk

Abstract

This paper presents a multimodal corpus of comparable pack messages and the concordancer that has been built to query it. The design of the corpus and its annotation is introduced. This is followed by a description of the concordancer's interface, implementation and concordance display. Finally, some ideas for future work are outlined.

1 Introduction

This paper introduces a multimodal concordancer[1] that has been developed to investigate variation between messages on fast-moving consumer goods packaging from China, Taiwan and the UK. The need to develop such a concordancer arises from the fact that these *pack messages* are themselves multimodal. While they communicate through what Twyman (1985) calls the *visual channel*, messages are realized using a combination of three modes (*verbal*, *schematic*, *pictorial*). Moreover, the verbal components of visual messages are modulated and segmented through typography (Waller, 1987).

It is assumed that this multimodality will have complex implications for cross-linguistic variation within the genre of pack messages. The specific nature of these implications is not yet known, but variation in the construal of textual meaning and cohesion would seem to offer a good starting point for investigation. However, using purely linguistic annotation and a monomodal concordancer to analyze such material could reveal only part of the picture.

An existing annotation scheme, developed by the Genre and Multimodality (GeM) project[2], is well-suited to my needs. In addition to information about their verbal and visual realization, the scheme provides a mechanism for encoding the rhetorical relations between message components.

However, existing tools for multimodal analysis do not support simultaneous investigation of verbal, visual and rhetorical phenomena. While Baldry's (2004) multimodal concordancer supports multilayered analysis of video data, his approach does not support the segmentation of still visual layouts, let alone consideration of specific typographical realizations. From an altogether different perspective, the database developed as part of the Typographic Design for Children[3] project does allow access to such typographic information, but does not relate this directly to the linguistic realization of messages.

Their multimodal realization makes pack messages a rich testing ground for the new concordancer and Chinese and English offer great potential for looking at multimodal cross-linguistic variation. Typographic resources are constrained by the writing system of a given language: Chinese offers variety in reading directions and a consistent footprint for each character; English offers a range of case distinctions and a predictable reading direction.

2 Corpus design

I take each pack as a text: through the messages by which it is realized, it 'functions as a unity with respect to its environment' (Halliday and Hasan,

[1] http://corpus.leeds.ac.uk/~martin/

[2] http://www.purl.org/net/gem/

[3] http://www.kidstype.org/

Proceedings of the Linguistic Annotation Workshop, pages 57–60,
Prague, June 2007.

1976). In the corpus, each text constitutes a record. Each record consists of a set of files. These include the transcribed and annotated pack messages, and photographs of each pack face. In the future, pack metadata will be added to describe the product category to which the pack belongs, the product name, brand owner, variety and so on. I will also record the location and date of purchase of each sample. This will support query constraints at the level of the record (e.g. packs of a certain size) and will facilitate comparisons across time as well as across *locales*, or markets.

Packs are represented in the corpus in an unopened state. As far as possible, every message on each face of the pack which is visible in this state is recorded. There are good reasons for this. Sinclair (1991) makes the point that the differences across specific parts of a text may constitute regularity within a genre. In the context of investigation into cross-linguistic variation within a single genre, this observation seems particularly apt.

The selection of packs for inclusion in the corpus will be made in cooperation with an industrial partner. Packs will be selected from product categories in which the partner is active, or seeks to participate, in all three locales. A combination of popular local brands as well as locally established global brands will be selected. Thus the packs will be comparable commercially as well as in terms of the communicative functions that they perform.

3 Corpus annotation

The GeM scheme is described comprehensively by Henschel (2003). It implements stand-off annotation in four XML layers. The *base* layer segments the document. The resulting *base units* are cross-referenced by layers which describe *layout*, *rhetorical structure* and *navigation*.

Within the layout layer, there are three main sections: layout *segmentation* (each *layout unit* contains one or more base units), *realization* information and a description of the *layout structure* of the document. These components allow a comprehensive picture of the typographic realization of the messages to be built, from details such as font family and colour to information about the composition of each pack and the location, spacing and framing of *chunks* of layout units.

Rhetorical relations between annotated units are expressed in terms of Rhetorical Structure Theory (Mann and Thompson, 1987). In the GeM implementation, RST has been extended to accommodate the graphical elements found in multimodal texts. RST annotation provides a way to identify patterns in the construction of messages and to make comparisons across the corpus. It might be that more RST relations of a specific type, e.g. *elaboration*, are found in messages from a particular locale. Such observations might support or contest claims, such as that packs from developing markets conventionally carry more information about how to use the product. In combination with the layout layer it will also be possible to look for patterns in the choice of semiotic mode used to realize messages involving specific types of relation, such as *evidence*.

In sum, the aim of the annotation is not to support low-level lexicogrammatical analysis, but rather to facilitate the uncovering of patterns in the linguistic and typographical realization of pack messages and to relate these to semantic values expressed in terms of RST relations. Such patterns may reflect local design conventions and language-dependent strategies for ensuring textual cohesion.

So far annotation has begun with several UK and Taiwan packs. All annotation has been performed manually and has proved costly in terms of time. In future it is hoped that at least some annotations may be generated through the conversion of digital copies of designs obtained directly from brand owners.

The pilot annotations have identified a number of ways in which the GeM scheme will need to be extended to accommodate the genre of pack messages and important aspects of Chinese typography: the lists of colours and font families enumerated in the DTD are not sufficiently extensive or delicate and there is no mechanism in the layout annotation layer to record the orientation and reading direction of text.

4 The prototype concordancer

4.1 Design aims and system overview

The concordancer is an established tool for linguistic analysis. Concordance lines, which show instances of a key word in their immediate contexts,

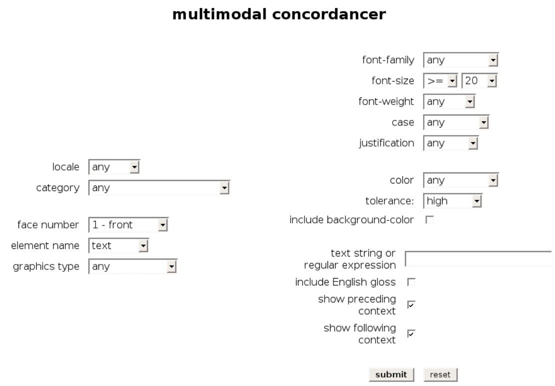

Figure 1: Multimodal concordancer interface

have proved useful in uncovering patterns of usage and variation that may not be apparent either from reading individual texts or from consulting reference resources, such as dictionaries and grammars.

My aim was to develop a similar tool to support multimodal analysis. Such a tool should be able to combine questions relating to the verbal components of messages with those relating to the typographic resources through which they are realized. It should do this in such a way that queries can easily be built and modified. To this end, a user interface is needed. Finally, the concordancer should be usable without the need for local installation of specialist client software.

In order to meet these requirements, I adopted a web-based client-server model. The user interface is shown in Figure 1. The concordancer is implemented in Perl as a CGI script. XPath expressions are used to identify matches from among the XML-annotated packs and to handle cross-references across annotation layers.

Using the concordancer interface to build a query is a process of moving from the general to the specific. By default, all constraints are relaxed: submitting a query with these selections will return every annotated message in the corpus. More usefully, selections can be made to constrain the set of records searched and the linguistic, typographic, and pictorial realization properties of messages to match.

4.2 Search criteria

The search criteria are grouped into high- and low-level selections. I will introduce the high-level selections first.

Locale and category selections control the set of records to be processed.

Given the notion of generic regularity in the differences between different parts of texts, it seemed sensible to allow queries to be constrained by pack face. Looking at the front of a shampoo bottle might be seen as akin to looking at the abstract of an academic paper. This is a step towards implementing more specific constraints about the on-pack position of messages. The pack face constraint, as with most of the remaining selections, is implemented in an XPath expression. The remaining high-level selections constrain the type of encoded element to include in the search.

The first group of low-level selections relate to specific font properties.

The colours used to realize messages are described in the corpus using hexadecimal RGB triplets. While this affords precision in annotation, it also means that some calculation is required to support searching. The current approach is to take any colour selected by the user from the menu and calculate the distance between this and the RGB value for each candidate match. If this distance falls within the tolerance specified by the user, the colour is considered to match. Thus a search for *green* may match RGB values representing various hues.

Finally, all matching layout units are cross-referenced with the base units that they realize. If the user specified a pattern to match (a string or regular expression), this is tested against the string value of the base unit.

4.3 Concordance display

The final options on the interface control the display of the resulting concordance. In the pilot annotations, an English gloss for each Chinese pack message is recorded as an XML comment. These glosses may be reproduced in the concordance. The other display options control whether to display the base unit preceding and/or following the match.

Figure 2 shows the results of a query generated from the selections shown in Figure 1. This is a search for verbal messages on the front of packs which are realized in a large font. Unsurprisingly, in each case, this returns the product name which is conventionally salient.

Details about the search query are given above the

Figure 2: Multimodal concordance example

concordance. Depending on the specific query, this may include selections for locale and product category, the XPath expression which identifies candidate layout realization units, the colour selection and the search string or regular expression.

Information relating to each match is then displayed. As in a traditional concordancer, matches are presented together with the context in which they are found. Optionally, this context includes the preceding and following base units. Moreover, the notion of context is extended to include the visual environment in which each match is found. The colour used on-pack to realize the matching message is reused in the presentation of the match. A thumbnail image of the pack face on which the match is found is also presented, as is information about the typographic realization of the match, taken from the layout annotation. Links are provided to high resolution photographs and to each annotation layer for the pack from which the match is retrieved.

The display of the thumbnail is a step towards a more specific indication of the position of each match on the pack. In the future, I hope to use information from the layout annotation to generate a visual representation of the layout chunk in which each match is found.

The number of matches found is given below the concordance.

5 Conclusions and future work

The prototype concordancer is rather slow: it takes just under a minute to process and print every unit in the pilot corpus and the time taken will increase as more packs are added. But it works. It has also been tested with files taken from the original GeM corpus. Once they have been renamed, following the conventions used by the concordancer, the legacy files integrate seamlessly into the new corpus.

As noted above, there is scope for further development in a number of areas. The pilot corpus needs to be populated with more packs. The GeM annotation scheme requires modification in certain details. It might also be useful to add an annotation layer to record translations of the string values of base units rather than using XML comments for this.

As for the concordancer, support for queries based on the rhetorical relations between message components is the next major step. Other planned functionality includes the generation of typographically realized layout chunks which contain query matches and the calculation of collocation statistics which may be compared across sets of records.

Finally, more work is needed to see whether the concordancer is useful for the kind of analytical work it has been developed to support.

References

Anthony P. Baldry. 2004. Phase and transition, type and instance: patterns in media texts seen through a multimodal concordancer. In Kay O'Halloran, editor, *Multimodal discourse analysis: Systemic-functional perspectives*. Continuum, London.

M.A.K. Halliday and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman, London.

Renate Henschel, 2003. *GeM Annotation Manual Version 2*. GeM Project.

William Mann and Sandra Annear Thompson. 1987. Rhetorical structure theory: A theory of text organization. Technical report, Information Sciences Institute, Los Angeles.

John Sinclair. 1991. *Corpus, concordance, collocation*. Oxford University Press, Oxford.

Michael Twyman. 1985. Using pictorial language: A discussion of the dimensions of the problem. In Thomas Walker and Robert Duffy, editors, *Designing Usable Texts*, chapter 11. Academic Press, Orlando, Florida.

Robert Waller. 1987. *The Typographic Contribution to Language*. Ph.D. thesis, University of Reading.