On the Analysis of Multimodal Documents at the Level of Pragmatics

John H. Connolly and Iain W. Phillips

Department of Computer Science,
Loughborough University,
Loughborough, LE11 3TU, UK,
J.H.Connolly@lboro.ac.uk, I.W.Phillips@lboro.ac.uk
WWW homepage: http://www.cs.lboro.ac.uk

Abstract. A genuine comprehension of the process of human communication within organisations requires a detailed understanding of pragmatics. Accordingly, our aim here is to work towards a richer means of analysing and describing multimodal documents at the pragmatic level than is currently available within the framework of Organisational Semiotics. Our basic strategy is to take six of the most widely known aspects of discourse analysis as practised in linguistics and attempt to extend them to multimodal documents. These aspects are: (i) discourse organisation, (ii) order of presentation, (iii) communicative acts, (iv) adjacency sequences, (v) rhetorical relationships and (vi) affect. We find that all six of these aspects of discourse analysis are, indeed, capable of being extended to the analysis of multimodal discourse, at least in terms of basic principle. This encouraging result provides us with a viable starting-point for the analysis and description of multimodal documents at the level of pragmatics.

Keywords: Discourse analysis, multimodal, multimedia, semiotics.

1 Introduction

Stamper [13] offers a six-level framework for the analysis and description of the signs used in communication. This framework recognises:

– The level of the social world.
– The level of pragmatics (concerned with the use of signs in context).
– The level of semantics (concerned with meaning).
– The level of syntactics (concerned with abstract form or structure).
– The level of empirics (concerned with observable patterns).
– The level of the physical world.

In the present paper we intend to focus attention upon one of these levels in particular, namely the level of pragmatics. Our aim is to work towards a richer means of analysing and describing multimodal documents, whether in the form of paper or of computer-based productions such as screen displays, than is currently available within
the Organisational Semiotics (OS) framework. Our main purpose, therefore, is to contribute to the theory of OS. However, the enhancements we propose may also be of some interest to designers of multimedia documents (such as web pages) who adopt OS as the basis of their approach.

As a starting point we may take note of the ways in which pragmatic-level analysis is carried out in linguistics, where it is reasonably well developed, and where it is centred around the notion of discourse. From the linguistic point of view, discourse consists in the process of communicating with an audience by means of coherently organised text (in the case where the visual medium of language is used) or talk (in the case of the auditory medium of language). The resulting text or talk is thus the product of the discourse process [2]. A discourse can take the form either of a monologue, where the author addresses the audience with no opportunity for these two participant-roles to be reversed, or of a dialogue, where two or more participants take turns at addressing the other(s).

Discourse Analysis (DA) involves the application of any of a number of concepts and methods for describing discourse or its products. Some of the main concepts associated with DA will be outlined in the ensuing sections.

Of course, linguistic DA is designed for the analysis of language, as distinct from non-linguistic modes of communication. On the other hand, in order to cope with the various kinds of linguistic and non-linguistic sign that can be found in multimodal discourse, it is necessary, as argued by Connolly and Phillips [4], to look to semiotics, which offers the required breadth of scope.

In fact, as Kress and van Leeuwen point out, discourse is, in general, not purely linguistic but multimodal in nature [7]. Talk is generally accompanied by non-verbal communication, such as gestures and facial expressions, while language-text is supplemented by the non-linguistic information conveyed through phenomena such as handwriting style or choice of typeface. Hence, it is quite natural to try to extend linguistic DA into the domain of multimodal communication in which language and other semiotic modes are combined.

There have been previous attempts to extend the application of linguistic concepts into the domain of semiotics, for instance O’Toole [10], Kress and van Leeuwen [7] and van Leeuwen [14]. In the present paper we wish to proceed along the same path. Our treatment of discourse will retain the text-focussed flavour of linguistic DA, rather than exploring the broader, socially-focussed approach found in other areas of the semiotic literature.

Of course, although linguistic approaches to discourse analysis may provide a starting point for multimodal DA, we must expect at least some of them to need modification when their coverage is extended. Multimodal documents manifest certain complications, compared with unimodal language-texts. Not only are they the product of the deployment of signs from more than one semiotic mode, but there is typically a degree of integration between one mode and another (for example between a text and an illustrative picture or diagram). Furthermore, not all modes of communication are as highly systematised as language [8]. Hence, the extension of linguistic DA to multimodal documents presents us with a genuine challenge.
In what follows, we shall consider six of the most widely known aspects of DA as practised in linguistics and attempt to extend them to multimodal documents. In order to keep the treatment within practicable bounds, we shall confine ourselves to documents combining language-text with images. These constitute the most common type of multimodal document used in contemporary organisations. From the point of view of semiotics, they typically contain a combination of both symbolic and iconic signs.

The six aspect of DA we shall consider are:

– Discourse organisation.
– Order of presentation.
– Communicative acts.
– Adjacency sequences.
– Rhetorical relationships.
– Affect.

These terms will be explicated as we proceed.

It is acknowledged that other approaches to discourse pragmatics are possible besides these. For instance, the concepts of reference, co-reference and deixis, as well as those of presupposition and implicature, have been explored by other authors concerned with pragmatics. Moreover, approaches focusing on sociolinguistic variation or ethnography have been developed. However, as far as the analysis of multimodal documents is concerned, we are content to leave those topics and perspectives as subjects for future research. Again, the application of multimodal DA to other forms of IT-supported communication, for instance contemporary mobile phones, would be interesting, but must wait until another occasion.

2 Discourse Organisation

2.1 In Unimodal Language-Based Communication

The first aspect of discourse that we shall consider is its overall organisation. Linguistic DA shows that both written and spoken discourse generally has some kind of hierarchical structure. In written texts, sentences are typically grouped into paragraphs, and in some kinds of document there are higher layers of organisation as well. For instance, a book or a report may be divided into chapters, sections, sub-sections, and so forth. Formal speech, too, tends to have a structure. Informal talk is not usually so rigidly organised, but usually it takes the form of dialogue, in which a hierarchical structure can readily be discerned. Dialogue is organised into units called turns, each of which represents a contribution from one or other of the participants to the conversation.

The smallest unit in a discourse is a communicative act. This may function as a statement or a question, for example, and typically consists of a single sentence. (The intricacies of the relationship between communicative acts, sentences, clauses and the prosodic units associated with intonation need not detain us here.) Employing terminology from Coulthard [5], communicative acts combine to form moves in the discourse. A move typically corresponds either to a paragraph or to a turn, depending on the type of discourse involved. Within a dialogue a sequence of moves on the same topic constitutes an exchange. An example of an exchange is as follows:
ANDREW: It’s warm in here. Shall I open the window?

BARBARA: That sounds like a good idea.

Within this exchange, Andrew makes the first move, which consists of two acts: a statement and then a question. Barbara’s reply constitutes the second move, and contains just one act, which is a statement.

It is possible for a text document to contain more than one more-or-less independent discourse structure. An example is a newspaper containing a number of stories with little or no interrelationship among them. Each of them represents a different strand of discourse.

An additional complication arises in hypertext documents. Unlike traditional printed materials, which are usually designed to be read from start to finish, hypertext is generally designed in such a way as to provide more than one equally coherent path through the document. Nevertheless, language-based hypertext normally exhibits some degree of hierarchical structure, at least at the local level, for instance paragraphs within an individual page.

2.2 In Multimodal Communication

Let us now consider discourse organisation in relation to multimodal documents consisting of language-text together with images. An immediate complication arises from the fact that some images constitute part of the discourse (for instance, a diagram representing the architecture of a computer system described in the text which it accompanies) whereas others do not (for example, pictures included purely for decoration). We shall use the term consolidated to describe images that form part of at least one strand of discourse within the document concerned. Images that are not consolidated will be termed perfunctory.

When a multimodal document is primarily based on language-text together with consolidated images that serve to illustrate or enrich it, we can expect to find that insofar as the text is organised into units such as chapters, sections and paragraphs, then this organisation will confer a basic hierarchical structure upon the whole document. The consolidated images will be attached to the language-text in some way. Often they will be specifically referenced, by means of some expression like ‘see Figure 1’, which we shall here call a pointer. In simple cases, if we represent the hierarchical structure of the text as a tree, some of whose nodes are pointers, then the images may be conceived of as leaf-nodes attached to those pointer-nodes. However, if the same image is referenced from more than one place in the text, then a pure tree-structure representation is no longer feasible. It is also possible for an image to be consolidated and yet not to be referenced by a pointer, as for instance when a :) is included in a sentence within an email message. As for perfunctory images, these are associated only loosely with the discourse and do not form part of its actual structure. In sum, then, the effect of adding images to language-text is, typically, to dilute the hierarchical structure of the document, but without dissolving it entirely.

Even in those multimodal documents that cannot be regarded as being fundamentally composed of text accompanied by illustrations, it is usually possible to discern some degree of hierarchy. For instance, consider a web page that consists mainly of
buttons providing links to other parts of the web site; some of the buttons have textual labels, but there are no paragraph-length stretches of text. If the buttons are grouped in some way, then there will be a hierarchical relationship between an individual button and the group of which it forms part. Moreover, the index-page of any web site with more than one page can be regarded as at the top level of a hierarchical structure.

A further consideration is that images within multimodal documents are themselves often multimodal internally. Buttons may have textual labels, pictures may have titles, diagrams may have both titles and labels, and so on. Such images are composite signs, partly iconic (in respect of what they portray) and partly symbolic (in particular, with regard to the textual elements just mentioned). They thus have an internal semiotic structure, which is hierarchical in nature.

3 Order of Presentation

3.1 In Unimodal Language-Based Communication

It has been remarked by many authors that sentences tend to be organised in such a way as to proceed from the known to the unknown [3]. For instance, consider the following two sentences:

I should like to introduce Mr James Smith. He is the British ambassador.

The first sentence opens with the word ‘I’, which refers to the person speaking, who is present to the audience before the utterance begins and can therefore be regarded as a ‘known’ entity from the outset. However, the purpose of this initial communicative act is to introduce Mr James Smith, whose identity is presented as constituting fresh information to the audience (whether they were genuinely ignorant of his name or not). The second sentence begins with the word ‘he’, which refers back to James Smith, who is by now a known entity, whereas the fact that he is the British ambassador is presented as being novel information. Thus, in both these sentences we see a progression from the already known, or given, to the previously unknown, or new.

The given-new progression may also be found to characterise larger sections of text than the individual sentence. For instance, the opening paragraph in a book may introduce a new concept. However, once it has been introduced, the concept becomes given information. Therefore, if a subsequent paragraph explains how the concept may be subclassified, then the subclassification is presented as new information about (what is by then) a known entity.

3.2 In Multimodal Communication

The idea of given and new material has also been applied to multimodal discourse. Kress and van Leeuwen point out that in many (though by no means all) visual compositions, the horizontal axis is employed to present a progression from given to new [7]. For instance, a diagram of the communication process might well be arranged as in figure 1.
The left-to-right flow not only represents temporal progression; it also reflects the status of the message, which is known to the sender but unknown to the recipient (at least until the latter has had the opportunity to perceive and interpret it).

Kress and van Leeuwen also claim that the vertical axis of visual compositions can likewise be significant [7]. In language-texts, as everyone knows, it is normal for a heading to be placed above the body of text to which it applies. Generally, the body of text gives further information about the topic adumbrated in the heading. Thus, we find a progression from the generality of the concept evoked by the surmounting heading to the particularity of the detail spelt out beneath. Kress and van Leeuwen describe this as a progression from the ideal (the quintessence of the concept) to the real (the less rarefied particulars). It can be found in some (though by no means all) multimodal documents, for instance taxonomy diagrams with the root of the tree at the top (rather at the bottom, which is what the analogy of a real tree would suggest), as in figure 2.

![Diagram of the communication process](image)

**Fig. 1.** A simple diagram of the communication process.

A third presentational dimension identified by Kress and van Leeuwen is that of centre and margin [7]. Again, this can be found in some (though by no means all) visual compositions. For instance, in a (labelled and hence multimodal) chart depicting the variation of pressure within a vessel against time, the curve of the graph occupies centre-stage, while the axes representing time and pressure are relatively peripheral.
4 Communicative Acts

4.1 In Unimodal Language-Based Communication

The communicative acts, or *illocutionary acts*, into which discourse may be analysed can be subdivided into categories, such as the following proposed by Searle [12]:

- **Representatives.** These commit the speaker or writer to the veracity of what he or she is saying or writing.
  Example: ‘The report has been completed.’
- **Directives.** These are produced by the speaker or writer in order to try to cause the addressee to do something (which could be simply to provide an answer to a question).
  Examples: (i) ‘Please sit down.’ (ii) ‘What is your full name?’
- **Commissives.** These commit the speaker or writer to do something in the future.
  Example: ‘I shall inform you of our decision within 48 hours.’
- **Expressives.** These serve to externalise a psychological state.
  Example: ‘Thank you for your good wishes.’
- **Declarations.** These, when uttered by a duly authorised person, serve to alter the official state-of-affairs.
  Example: ‘I declare the meeting closed.’

An alternative classification is offered by Halliday [6], who distinguishes among the following:

- Giving information.
  Example: ‘It is cold.’
- Giving goods/services.
  Example: ‘Have a chocolate!’
- Demanding information.
  Example: ‘When does the train for Lisbon leave?’
- Demanding goods/services.
  Example: ‘Pass the disk, please.’

4.2 In Multimodal Communication

Any of the communicative acts that can occur in unimodal language-based communication can, of course, also appear in the language-text component of a multimodal document. What is more interesting, however, is the suggestion by Kress and van Leeuwen [7] that the concept of communicative acts can be applied to images as well as to language-text. Using Halliday’s classification, they propose that images can either (i) give information (their usual function) or (ii) demand a specific kind of service, namely that the viewer should enter into an imaginary social relationship with a person portrayed in the image. For instance, supposed that an advertisement shows an attractive young adult smiling towards the viewer. The idea here is that the viewer should mentally engage with and be mildly charmed by the young person, and should associate these positive connotations with the product being advertised.
If we employ Searle’s terminology rather than Halliday’s, then we are offered a finer-grained classification for the type of image that offers information. For instance, a colour photograph of the moon during an eclipse can contribute a representative act to a multimodal discourse if it communicates the appearance of the moon during such an occurrence. Or a drawing of a skull and crossbones warning of mortal danger may be taken as a commissive act, promising that anyone who dares to step forward into the territory guarded by the sign will be shot dead or blown to pieces. Or a picture of a sad-faced person dressed in black (or other appropriate colour, depending on the culture) may serve the expressive function of indicating mourning. As for images that demand a service, these would be classed as directives in Searle’s classification.

Given that the same classification can be applied in this way not only to language-text but also to pictures and drawings, clearly it can also be applied to multimodal combinations of the two. For instance, a picture of the highest mountain in Wales accompanied by the title ‘Mount Snowdon’ is a communicative act of Searle’s representative class.

5 Adjacency Sequences

5.1 In Unimodal Language-Based Communication

In interactive discourse, or dialogue, it is often possible to discern coherent patterns involving a succession of communicative acts. For instance, a question is typically followed by an answer, which may be followed by a reaction from the questioner, as in the following exchange:

\begin{verbatim}
ANDREW: Did you see Carol today?
BARBARA: Yes.
ANDREW: Good!
\end{verbatim}

A pattern such as this involving a series of communicative acts, one after the other, is an adjacency sequence. Many such sequences involve just a pair of acts, for instance a statement followed by an acknowledgement, as in the following exchange:

\begin{verbatim}
ANDREW: I’m going to buy a paper.
BARBARA: OK.
\end{verbatim}

This two-term sequence is an adjacency pair \cite{11}.

Of course, there is no guarantee that Barbara will respond to Andrew’s statement with a compliant acknowledgement (the expected response). Instead, she may respond in some non-expected way, for instance by saying ‘No, you’ve got to stand here and look after the luggage!’; thus making a contradictory statement to Andrew’s.

Halliday proposes the following classification of adjacency pairs, based on his categorisation of communicative acts \cite{6}:

- When giving information: Statement followed by acknowledgement (expected response) or contradiction (alternative response).
– When giving goods/services: Offer followed by acceptance (expected response) or rejection (alternative response).
– When demanding information: Question followed by answer (expected response) or disclaimer (alternative response). An example of a disclaimer would be ‘I don’t know’.
– When demanding goods/services: Command followed by undertaking (expected response) or refusal (alternative response).

Note that by no means every communicative act in a discourse is part of an adjacency sequence. Nevertheless, adjacency sequences constitute an important part of conversation.

5.2 In Multimodal Communication

Traditionally, documents have had the character of monologues rather than dialogues. However, present-day technology affords the possibility of interactive written documents. For instance, clicking on a link within a web page sometimes results in the appearance of a dialogue box with a message along the lines of ‘This item requires a plug-in. Would you like to download it?’. The user is then expected to click either on the button labelled ‘Yes’ or on the button labelled ‘No’. This dialogue-exchange is multimodal (employing text and buttons) and constitutes an adjacency pair, comprising an offer followed by either an acceptance (clicking on the ‘Yes’ button) or a rejection (clicking on the ‘No’ button). This kind of exchange is very widely used in interactions between humans and computers.

6 Rhetorical Relationships

6.1 In Unimodal Language-Based Communication

Another aspect of discourse structure, applying to both monologue and dialogue, lies in the rhetorical relationships between communicative acts. These relationships serve the purpose of progressing the discourse in a coherent manner and are essential to the development of a convincing argument. They form the basis of an approach known as Rhetorical Structure Theory (RST), introduced by Mann and Thompson [9].

In RST a number of rhetorical relationships are recognised. Some examples are as follows:

– Rhetorical relationship of justification.
  Example: ‘I’ll have to postpone our meeting. I’ve been told to go to London tomorrow.’ (The second statement is included in order to justify the first.)
– Rhetorical relationship of exemplification.
  Example: ‘There have been a number of problems recently. For instance, the electricians have threatened to go on strike.’
– Rhetorical relationship of elaboration.
  Example: ‘We need to move to larger offices. There are some suitable ones available in South London at the moment.’ (The second statement provides further information relevant to the first, but does not, for instance, exemplify or justify the need to move.)
– Rhetorical relationship of enumeration.

Example: ‘We need to write two letters. Firstly, we must reply to James Smith. Secondly, we had better ask the bank for an increased overdraft facility.’ (The two non-initial sentences in this extract enumerate the tasks noted in the opening sentence.)

Verschueren lists more relationships of this kind [15].

6.2 In Multimodal Communication

As with communicative acts, any rhetorical relationship that can occur in unimodal language-based discourse can also, obviously, appear within the language-text component of a multimodal document. Interestingly, however, some (though by no means all) of the rhetorical relationships found inside language-text can also exist between a unit of language and an image. For instance, consider a text, which contains the sentence ‘The building contains twenty furnished offices’, illustrated with a picture of one of the offices concerned. The image thus provides an example of the offices mentioned in the quoted sentence. In other words, the rhetorical relationship of exemplification connects a linguistic and a non-linguistic item.

In other cases, an image may provide further information about an entity mentioned in a piece of language-text, but not actually provide an example of it. For instance, if a unique name is mentioned and a photograph of the named object or person placed alongside, then the image serves as an elaboration of the linguistic item (the name). To take an actual example, suppose that on some web page, the name ‘Tower Bridge, London’ is mentioned, along with a photograph of the bridge in question. Here, the picture does not furnish an example of Tower Bridge, London, as there exists only one such object. Rather, it elaborates the textual allusion by providing additional, visual information as to the appearance of that bridge.

The relationship between language and images is discussed by Barthes [1], who speaks of language-text providing an image with anchorage. Typically, an individual image can be interpreted in more than one possible way; and what an associated piece of language-text can accomplish is to fix, or anchor onto, one of the available interpretations. For instance, suppose that a picture of the centre of a particular town shows a large square, at the far side of which, and occupying the centre of the picture, is the town hall. Is this picture supposed to be a view of the town hall or of the central square in general? Clearly, either is interpretation is possible. However, if the picture is accompanied by the title ‘The Town Hall’ (thus forming a composite sign), then the intended interpretation is fixed or anchored. It would seem, then, that insofar as it serves to progress the multimodal discourse in a coherent manner (from the reader’s point of view), anchorage can fairly be regarded as a kind of rhetorical relationship, which can hold between linguistic and non-linguistic parts of a composite sign.
7 Affect

7.1 In Unimodal Language-Based Communication

A further aspect of discourse lies in the attitude expressed by the author towards the subject-matter. Attitude can be communicated by various linguistic means, including the following:

– Through the connotations of individual expressions.
  Example: the word ‘plagiarism’ carries pejorative connotations.
– Through the intonation of spoken utterances.
  Example: if someone is told about an unfortunate event and replies by saying ‘Tragic!’ on a low monotone, then this is likely to be perceived as expressing an ironic attitude.
– Through the style of language employed.
  Example: a scientific style is associated with an attitude of objectivity towards what is being described.

Attitude conveyed by these or other semiotic devices is termed affect.

7.2 In Multimodal Communication

Affect can also be very effectively communicated through multimodal discourse. The text in a document can be illustrated with pictures that convey a happy mood, for instance to show travel destinations in a good light in order to attract tourists. Alternatively, pictures showing the aftermath of natural disasters like earthquakes and hurricanes can evoke reactions such as shock and sympathy. Images of this kind often have an impact that could almost certainly not be achieved through words alone.

8 Conclusion

In this paper we have taken six aspects of discourse analysis that originated in linguistics and considered their possible extension to the analysis of multimodal discourse. In the event, all them have proved capable of being extended in this way, at least in terms of basic principle. Admittedly, in some cases the extension has been only partial (for instance, in the case of communicative acts). Nevertheless, considering that the linguistic DA methods were not developed with multimodal discourse in mind, their capacity for extension into this broader domain has turned out to be surprisingly good. The result happily provides us with a viable starting-point for the analysis and description of multimodal documents at the level of pragmatics.

This proposal of a method for the pragmatic analysis of multimodal documents constitutes the essence of the present paper’s contribution to the theoretical armoury of OS. Given that multimodal documents are instruments of communication, and communication is of fundamental concern within semiotics, the application of DA as an aid to our understanding of these matters offers potentially important benefits.
It is hoped that future work will lead to further developments along the same lines, for the purpose of enhancing the capacity of the OS framework to accommodate the richness of multimodal communication. Further research could usefully include an attempt to formalise the DA of multimodal communication, with a view to enhancing our understanding of how best to provide support for multimodal human-computer interaction.

References