

Andrew Salway and Chris Frehen

“Words for Pictures: analysing a corpus of art texts”

Procs. International Conference for Terminology and Knowledge Engineering, TKE 2002.

Words for Pictures: analysing a corpus of art texts

Andrew Salway and Chris Frehen

Department of Computing, University of Surrey
Guildford, Surrey GU2 7XH
United Kingdom
a.salway@surrey.ac.uk

Abstract

The use of *collateral* text for indexing images is a promising approach for closing the ‘semantic gap’ between image databases and their users, as demonstrated by systems which process HTML tags and newspaper photograph captions to extract metadata for images. This approach can be extended in specialist domains like fine art in which the ‘semantics’ of paintings are articulated in the extensive discourses of subject experts. Thus we may talk of a *corpus* of collateral texts, gathered from the Web, to provide a conceptual framework for the organization of visual information through the extraction of terminology and knowledge-bases. The premise is that experts systematically articulate the subject matter of paintings and other domain knowledge using a *special language*. This paper analyses a corpus of art texts, specifically painting captions and artist biographies, to examine evidence for a special language of art and to evaluate the use of such a corpus for organising and accessing visual information in an on-line art gallery.

1. Introduction

A painting in an art gallery is usually accompanied by a caption that gives salient information about the painting and its artist to help the viewer/reader appreciate the work of art. In digital libraries, where the physical organization of information need not impinge on its ‘intellectual’ organization and presentation, it is possible for a painting to be accompanied by many *collateral* texts. From a caption the user may bring up a biography of the artist, critiques of the work by art scholars, textbook explanations of the techniques used, histories of events that inspired the painting, and so on. As well as being available to the user of an on-line art gallery these texts could be used by a digital library system as a source of terminology for organizing visual information.

Users can access material in digital libraries in two ways: either by submitting queries for which the system returns a list of texts, images, videos, etc. which best match the query; or by browsing through a collection of multimedia artefacts via hypermedia links. It is important to note that images are multi-faceted and may be organized according to a range of attributes. For example paintings could be indexed according to artist, date, type of painting (e.g. ‘landscape’, ‘portrait’), artistic movement (e.g. ‘impression’, ‘expressionism’) and even according to what inspired the artist to paint them. Access to the paintings could be enhanced with a knowledge-base of facts about artists and artistic movements to expand queries and automatically instantiate hypermedia links.

Of particular importance for the intellectual organization of a digital art gallery is the subject matter of the paintings. It has been noted from studies of user queries that “multiple levels of attributes are needed, ranging from the specific item named, to the generic category of the item, to the item’s meaning” (Jørgensen, 1999). These levels reflect those described by Panofsky’s theory of meaning in art (1939), and later used as the basis of a framework for indexing pictorial information by Shatford (1986). The levels are: *pre-iconography* – a

generic description of the objects and actions represented in the picture; *iconography* – identification of specific objects and events based on cultural knowledge; and, *iconology* – an interpretation of meaning requiring knowledge about the artistic, social and cultural setting to which a picture belongs.

The concern with subject matter in indexing visual information mirrors a general concern for the developers of multimedia information systems with *content*. Santini (2001) notes a “semantic gap” between the planes in which users and image databases operate: users of an on-line art gallery may not typically conceive their information needs in terms of the colours and shapes of the paintings in which they are interested. Rather they are more likely to be interested in the *semantic content* of images, like the subject matter of paintings, and other domain specific attributes, like scholarly classifications of paintings by artistic movement.

The *QBIC* system (*Query By Image Content*) has been deployed by the Hermitage Museum, an on-line art gallery, to allow users to make visual queries, i.e. to submit either sketches or other images for the system to return images with similar colours and shapes (Flickner, Sawhney, Niblack, et al., 1995). However other aspects of images are currently indexed manually with keywords and indeed several large and sophisticated frameworks have been developed for this purpose, e.g. the Library of Congress Thesaurus of Graphic Materials, the Art & Architecture Thesaurus and Iconclass. An alternative to manual annotation is to process collateral textual information that is already associated with images to produce indexing words. The *WebSEEK* system, which has indexed hundreds of thousands of images and videos on the WWW, selects keywords from the text of hyperlinks to images and videos on WWW-pages (Chang and Smith, 1997).

This kind of approach can be extended in specialist domains like fine art in which the ‘semantics’ of paintings are articulated in the discourses of subject experts so there is an extensive range of textual information that could be exploited for indexing images. This situation is

characterized in Figure 1 which shows how an image retrieval system can be grounded in the work of art scholars who analyse paintings and articulate their thoughts in texts like painting captions, artist biographies, textbooks, monographs and journal articles; the figure is based on a framework originally used in the development of a video annotation system for dance (Salway 1999).

Here we envisage a digital library in which an image database of paintings is integrated with a collateral text corpus. Some texts, e.g. painting captions, are linked specifically with one image such that keywords and perhaps richer representations of meaning may be extracted from the text to index the image. Other texts are used as a source for domain terminology and a knowledge-base, e.g. artist biographies and textbooks. The premise of this approach is that the subject matter of paintings, and other knowledge about art, is systematically articulated in experts' texts using a *special language*.

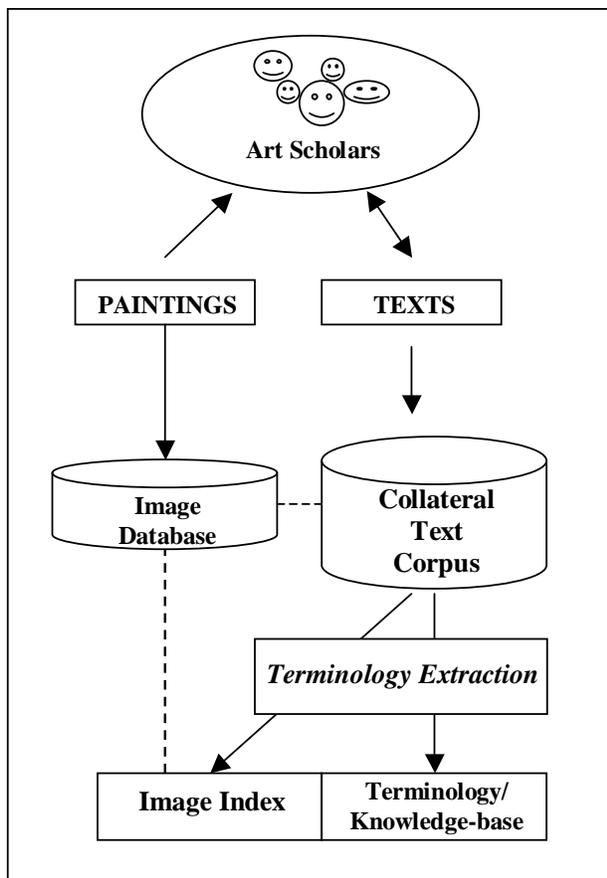


Figure 1. Using a collateral text corpus to facilitate access to images in a digital library of art.

2. Analysing a Corpus of Art Texts

Previously it has been observed that within specialist communities the spoken and written communications of experts lead to the development of a special language with an idiosyncratic vocabulary and grammar: in particularly special languages tend to have a profusion of terms with which to articulate the important concepts of a domain, and a restricted grammar to minimise ambiguity and ensure precise communication of ideas (Ahmad and

Rogers, 2001). Special languages have been investigated using the method of 'linguistic variance' whereby a corpus of specialist texts is compared statistically with a general language corpus. Research at the University of Surrey has shown how the method can be used to elaborate the lexicogrammar of special languages in domains including automotive engineering, nuclear physics and linguistics and how a range of corpus linguistics techniques can be used to automatically extract terminology and knowledge-bases from specialist corpora (Ahmad and Rogers, 2001; Ahmad, 2001). Some of these techniques are applied here to examine evidence for a special language of art, and to evaluate the use of a collateral text corpus for indexing images and building a knowledge-base in a digital library of art.

The main corpus used in this work comprises painting captions and artist biographies totalling about 800,000 words; see Table 1 for a breakdown. All these texts came from the Tate Collection – an on-line database of 25,000 works of art corresponding to works held by several Tate galleries. Some results are also presented from earlier work that was carried out before the current corpus was available. This work was based on corpora gathered from five on-line art galleries: further details are given at appropriate points in the paper. Corpus analysis was done using the freely available *System Quirk* (www.computing.surrey.ac.uk/SystemQ).

	Texts	Words
Captions	6715	691,121
Biographies	466	113,818
TOTAL	7181	804,939

Table 1: The main corpus of art texts analysed in this paper, comprising painting captions and artist biographies

3. The Specialist Terminology of Art

The 100 most frequent words in the corpus included 23 open-class words: this compares with two open-class words in the top 100 of the general language British National Corpus (BNC), and between 30-40 in other specialist corpora previously studied. Many of these words refer to works of art in general, e.g. *work(s)*, *painting(s)*, *art*, *artist(s)*, *paint(ed)*. The terms *landscape* and *portrait* make a broad distinction between two types of painting; and *colour*, *figures* and *forms* indicate the focus of a critic's attention when appreciating a work. For indexing purposes the only directly applicable terms are *landscape* and *portrait*: however, some of the more frequent terms are productive in the formation of compounds which will be useful labels, e.g. *abstract art*, *surrealist art*, etc. Words like *colour*, *shape* and *forms* suggest attributes by which paintings could be indexed. In the next 200 most frequent words the open-class words indicate further important concepts for the appreciation of art: *elements*, *shapes*, *theme*, *light*, *lines* and *composition* are further attributes of a painting; *century* and *period* highlight the importance of the history of art. Some words appear that might label broad categories of art, e.g. *abstract*, *natural*, *classical* and *modern*, and some everyday words appear which may indicate the subject matter of paintings in rather general terms, e.g. *sea*, *city*, *body*, *head*, *hand*, *house* and *room*. Finally, the words

inspired and *influenced* are frequent perhaps indicating the interest of art scholars in connecting works/artists with other works/artists and significant events in their lives.

In order to extract more specialist terminology the relative frequency of each word in the corpus was divided by its relative frequency in the general language BNC to give a *weirdness* value; a value close to 1 indicates 'normal' usage of the word in the special language corpus – a high value indicates unusually high usage and a likely specialist term. A list of 331 words was generated from the caption corpus by using the following thresholds: frequency > 5 and weirdness > 50 (words that did not appear in the BNC were not included). Just under half the words in the list were proper nouns – mostly artists' names with a few towns and countries. The rest of the words in the list for the most part referred to: art movements, e.g. *vorticism, surrealism, suprematism, expressionist, modernist, cubism*; techniques, e.g. *brushwork, brushstrokes, mezzotint, overpainting, portraiture*; types of work, e.g. *watercolour, printmaking, sketches, reliefs, triptych*; and, features of a painting, e.g. *monochromatic, figurative, naturalistic*. There were a small number of words which appear to refer to the subject matter of some of the paintings, e.g. *crucifixion, Daedalus, whaling*.

A list of candidate compound terms automatically extracted using System Quirk's *Ferret* module had a similar composition, i.e. a great many artist names, as well as art movements e.g., *surrealist group, pop art, abstract art, abstract expressionism, renaissance art*; types of work e.g., *pencil sketch, ink drawings, group portraits*; features of a painting e.g., *geometric abstraction*; and subject matter e.g., *biblical subjects, autobiographical resonance*. As might be expected many of the compound terms refer to sub-types of some of the more common single terms, e.g. *pop art, pencil sketch* and *group portrait*.

The frequency lists for the whole corpus give an insight into the conceptual framework used by scholars when writing about works of art, and indicate some general terms that may be used for indexing, or which refer to attributes that a work of art could be indexed on. Whilst this analysis gives an overview it is important to recognize that art scholars may employ different conceptual frameworks when analysing different kinds of paintings, such as abstract art and figurative art. Two sub-corpora were created of approximately 20,000 words each: one comprised captions for paintings exemplifying abstract art, the other for figurative art. Table 2 lists some terms that are more frequent in one of the sub-corpora than the other. When writing about abstract art there appears to be more emphasis on *colour, form* and *technique*: in contrast in writing about figurative art *mood, feeling* and *representation* are important.

Words used more in captions about abstract painting	Freq. Abstr.	Freq. Fig.	Abstr./ Fig. ratio
<i>structure</i>	15	0	INF
<i>abstract</i>	101	2	50.5
<i>element</i>	16	1	16.0
<i>shape</i>	26	2	13.0
<i>dynamic</i>	7	1	7.0
<i>technique</i>	14	3	4.7
<i>color/colour</i>	60	20	3.0
<i>line</i>	33	11	3.0
<i>style</i>	53	18	2.9
<i>form</i>	62	23	2.7

Table 2a: Some words appearing more frequently in captions for abstract art (Abstr.) than in captions for figurative art (Fig.), e.g. the word *element* occurs 16 times more often in the captions for abstract art.

Words used more in captions about figurative painting	Freq. Fig.	Freq. Abstr.	Fig./ Abstr. ratio
<i>mood</i>	3	0	INF
<i>place</i>	11	1	11.0
<i>portrait</i>	63	7	9.0
<i>scene</i>	27	3	9.0
<i>feeling</i>	6	1	6.0
<i>represent</i>	22	8	2.8
<i>sense</i>	10	4	2.5
<i>landscape</i>	31	17	1.8
<i>light</i>	15	9	1.7

Table 2b: Some words appearing more frequently in captions for figurative art (Fig.) than in captions for abstract art (Abstr.), e.g. the word *scene* occurs 9 times more often in the captions for abstract art.

4. Identifying a Painting's Subject Matter

The terms identified in the previous section say little about the subject matter of the paintings, i.e. the people, objects, events and themes perceived by a human viewer, though of course the captions are filled with words which describe the subject matter of paintings. The challenge is firstly to ascertain which words refer to subject matter – the captions are quite discursive and discuss other topics such as the artist and historical details. Secondly, it may be desirable to index paintings according to different levels of subject matter, e.g. the three levels proposed by Panofsky and used by Shatford. In this section we look in more detail at how experts write about paintings, and in particular at how they elucidate their subject matter. We take as a starting point some of the words that were noted as being frequent in the corpus: these words, like *depict* and *convey* appear to be potential 'cues' for keywords and larger text fragments which describe the subject matter of paintings in captions.

4.1. Cues for Different Levels of Subject Matter

The verb *depict* occurred 295 times in the caption texts (*depicts* – 207; *depict* – 79; *depicted (by)* – 9); with 119 occurrences the verb *convey* was less frequent but still prominent in the same texts (*convey* – 70; *conveys* – 36; *conveyed (by)* – 13). Some concordances of these verbs suggest they are used in distinct ways to elucidate the subject matter of paintings. In the following examples, Table 3, the grammatical subject of *depict* is always a work of art and what is ‘depicted’ relates, in Panofsky’s terms, to either the pre-iconographical level, e.g. *glass, pears, box, a group [of people], a riotous and debauched domestic gathering*, or the iconographical level, e.g. *a scene from the Trojan wars, a tram stop outside the Estacion di Francia*. Whilst what is depicted is usually described with nominals, there are examples of actions being described, e.g. *two women eating seafood*.

Example concordances of <i>depict</i>
this painting <u>depicts</u> a glass, two pears and a box
this work <u>depicts</u> a group struggling in a wind
a work which <u>depicts</u> a riotous and debauched domestic gathering
it <u>depicts</u> a scene from the Trojan wars
this work <u>depicts</u> a tram stop outside the Estacion di Francia railway station
this work <u>depicts</u> two women eating seafood at the famous Parisian restaurant Prenier

Table 3: Some concordances of *depict* which show how it is used to introduce pre-iconographic and iconographic subject matter.

The subject of the verb *convey* in its concordances, Table 4, is sometimes a work of art but there are also examples where a particular feature of the work is identified as doing the conveying, e.g. *[its] directness and simplicity, vivid network of lines and gestural brushwork, expressive use of colour and shape*. What is being conveyed, according to these examples, relates to Panofsky’s iconological level of subject matter and is described with abstract nouns, e.g. *claustrophobia, innocence, energy and movement*. In some cases the abstract noun is part of a phrase, e.g. *a sense of ___* or *a feeling of ___*.

Example concordances of <i>convey</i>
this painting has a directness and simplicity that <u>conveys</u> a feeling of innocence
using a vivid network of lines and gestural brushwork, Auerbach <u>conveys</u> a powerful sense of energy and movement
rustic simplicity is <u>conveyed</u> by her lack of jewellery
this composition <u>conveys</u> the claustrophobia of the interior of an omnibus
an expressive use of colour and shape to <u>convey</u> the subject’s character and mood

Table 4: Some concordances of *convey* which show how it is used to introduce iconological subject matter.

Before the current corpus was available one of the authors (who has received six years of tuition in art criticism) classified the usage of *depict* and *convey* in a different corpus of painting captions (305,913 words gathered from five on-line art galleries). The results confirm what the examples here suggest, i.e. that there is a regular preference by art experts to use *depict* when introducing pre-iconographical or iconographical subject matter, and to use *convey* to introduce iconological subject matter, Table 5.

Level of Subject Matter	Verb Used to Introduce Subject Matter	
	<i>depict</i>	<i>convey</i>
Pre-iconographical	81	0
Iconographical	59	4
Iconological	5	43
TOTAL	145	43

Table 5: Showing how often *depict* and *convey* were used to introduce different levels of subject matter in a corpus of painting captions (305,913 words).

The use of *depict* and *convey* was further examined by analysing the left-hand and right-hand contexts of the verbs in the 3rd person singular form, such that the left-hand context should include the subject of the verb and the right-hand context should include what it is that is being depicted / conveyed. The frequency lists that were generated support the claims that *depict* and *convey* are used in distinct ways. The left-hand context of *depict* was dominated by nouns that referred to the whole of the work of art, e.g. *painting, work* and *portrait*: for *convey* it was mostly nouns referring to aspects of paintings which rose to the top, e.g. *colours, elements* and *surfaces*. The top words for the right-hand context of *depict* included *movement, scene, landscape* and *Christ*: the corresponding words for *convey* were *sense, feeling, essence, mood* and *spiritual*.

If *depict* and *convey* are indeed used consistently to mention different aspects of a painting’s subject matter then they could be used as ‘cues’ by an image annotation system processing collateral texts. Based on the analyses presented here it would seem feasible for a system to associate keywords either with the pre-iconographic/iconographic levels or with the iconological level based on their co-occurrence with one or other of the cues.

4.2. Further cues?

A look at the list of highly frequent words in the corpus gives some other potential cues for elaborating the subject matter of paintings. These words may be loosely synonymous with *depict*, e.g. *allude, show, emphasise* and with *convey*, e.g. *evoke, express, symbolize*. The example concordances given in Table 6 suggest that they could also be exploited by a digital library’s indexing system to attribute keywords to different levels of subject matter.

Potential Cue	Frequency	Example
Similar to <i>depict</i> ?		
<i>allude</i>		pastels which <u>allude</u> to the death of Ophelia
<i>show</i>		here Rosetti <u>shows</u> Eve taking the fatal fruit
<i>emphasise</i>		they <u>emphasise</u> her rich costume and jewellery
Similar to <i>convey</i> ?		
<i>evoke</i>		The picture's harsh colours and lines <u>evoke</u> the discordant vitality of the modern city
<i>express</i>		The work also <u>expresses</u> a certain austere dignity
<i>symbolise</i>		This motif in Picasso's work <u>symbolises</u> the resistance of the Spanish popular front

Table 6: Other potential cues for elaborating the subject matter of paintings at different levels of meaning (see previous discussion of *depict* and *convey*).

5. Charting the History of Art

Scholars often study the work of artists in the context of the history of art, i.e. they group artists into 'schools' and 'movements' and trace connections, or influences, between particular artists and/or between particular works. Scholars also situate works of art in a more general historical and social context in order to explain particular events which prompted, or inspired, an artist to produce a specific work. The verbs *influence* and *inspire* occurred 403 times and 442 times respectively in the corpus of art texts. Here we analyse concordances of these verbs and discuss the potential for automatically building a knowledge-base of the history of art and its use in a digital library.

The two verbs occurred mainly in the passive, i.e. *influenced by* (80% of occurrences) and *inspired by* (79% of occurrences). As might be expected these phrases were used to different extents in the painting captions and the artist biographies: *influenced by* appeared four times relatively more frequently in the artist biographies than in the captions; *inspired by* appeared almost twice as often in the captions. It looks like that in writing an artist's biography an art scholar is more concerned with tracing influences on the artist, and that in a painting caption the focus switches to particular inspirations for the work. The examples in Table 7 exemplify this kind of usage.

Example concordances of <i>influenced by</i>
his paintings of the Thames were <u>influenced by</u> Whistler
his early works, which were <u>influenced by</u> Delacroix and Manet
where he was <u>influenced by</u> expressionism
Example concordances of <i>inspired by</i>
it was <u>inspired by</u> a view of the bay at St Ives in Cornwall
this picture was <u>inspired by</u> a performance of Shakespeare's play Macbeth
Severini was <u>inspired by</u> modern machinery

Table 7: Some concordances which show how *influenced by* and *inspired by* are used respectively to trace relationships between artists and artistic movements, and to specify the motivation for particular works.

The question arises, is there enough regularity in the usage of *influenced by* and *inspired by* for a machine to automatically extract facts about who and what influenced and inspired artists and paintings; in other words, can these phrases be used as knowledge elicitation cues? Working with an earlier corpus we analysed the left- and right-contexts of the two phrases using the following classification scheme:

PERSON (P): a proper noun (artist's name), personal pronoun or possessive pronoun (e.g. her work);

MOVEMENT (M): a singular noun ("expressionism"), a plural noun ("the expressionists") or a looser grouping of works ("18th Century Painting");

WORK (W): a specific work referred to by title ("Sunflowers") or by deictic reference ("this work"), or a particular aspect of a work ("the pose of the female figure");

ENVIRONMENT / WORLD (E/W): the local environment of the artist ("the plant outside his studio door") or larger-scale happenings ("political events", "scientific theories");

COMBINATION (C) & OTHER (O).

As Table 8a shows, the left-context of *influenced by*, that is the thing being influenced, was most often a person (50%); 40% of the time the subject of the verb was beyond the sentence - maybe because in a biography the subject can be taken as read. The right-context of *influenced by*, that is what was exerting the influence, was most often a PERSON (50%), followed by MOVEMENT (22%). The phrase *inspired by* was used about equally to refer to an artist or to a work (37% and 31%), Table 8b. Note that the prominent source of inspiration comes from the Environment/World class (40%). Given a list of artists' names and techniques for pronoun resolution it seems relatively straightforward to extract facts about which artists and artistic movements have influenced other artists in the history of art; a network of such influences could support hypermedia browsing through a collection of paintings. Though machine-based representations of inspirations might be harder to achieve, the cue 'inspired by' could at least be used to direct the user of an on-line gallery to pieces of expository text.

Left Context	Right Context						Total
	P	M	W	E/W	C	O	
PERSON	27%	11%	0	7%	5%	0	50%
WORK	1%	0	0	0	1%	0	2%
Beyond Sentence	22%	11%	0	0	7%	0	40%
Other	0	0	0	0	0	8%	8%
TOTAL	50%	22%	0	7%	13%	8%	100%

Table 8a: Usage of the phrase *influenced by* in a corpus of captions and biographies containing 258 instances of the phrase. The left and right contexts are classified according to the scheme described above.

Left Context	Right Context						Total
	P	M	W	E/W	C	O	
PERSON	16%	5%	0	14%	2%	0	37%
WORK	11%	2%	3%	14%	0	1%	31%
Beyond Sentence	3%	1%	0	10%	1%	0	15%
Other	2%	2%	1%	2%	0	10%	17%
TOTAL	32%	10%	4%	40%	3%	11%	100%

Table 8b: The same classification for 132 instances of *inspired by*.

6. Closing Remarks

For some, the link between a special language and the specialist knowledge of a domain is such that it is considered feasible to study the development of conceptual frameworks by observing linguistic changes, e.g. new vocabulary and new morphological inflections and derivations. In a 'multimedia' scenario, such as fine art, maybe there is a three-way link between language, knowledge and visual information. Such a link is important for the development of multimedia computing systems because the structure of a special language could be used to organise and index otherwise unstructured multimedia data like image and video files.

The link between language, vision and knowledge is of interest to computing scientists, particularly those concerned with artificial intelligence. The development of intelligent multimedia information systems benefits from the structure imposed by special languages in specialist domains of human expertise. The interest in special languages, studied for example through the automatic analysis of electronic text corpora, can perhaps be viewed as the practical and empirical end of a continuum at the other end of which are philosophers of science and philosophers of aesthetics who ask interesting questions about what the structure of language says about scientific and aesthetic knowledge.

Fine art is a good domain to explore the three-way link between language, knowledge and visual information because of its ready supply of images and written discourses. Importantly these discourses are grounded in a conceptual framework that has been developed over

many decades, if not centuries, by art scholars. Our interest in fine art is theoretical in that we would like to understand more about how art experts articulate their knowledge about art using special language, and practical in that we are concerned with developing content technologies to assist in accessing digital libraries.

The analysis of a corpus of art texts reported in this paper provides some evidence for a special language of art – a profusion of terminology and a restricted set of lexicogrammatical patterns. For the development of digital libraries the terminology may be useful for determining both metadata attributes and values. The consistent use of cues like *depict* and *convey* would support the automatic annotation of paintings at different levels of subject matter, whilst cues such as *influenced by* and *inspired by* could be used to automatically extract facts for a knowledge-base of art.

7. References

- Ahmad, K., 2001. The Role of Specialist Terminology in Artificial Intelligence and Knowledge Acquisition. In S.-E. Wright and G. Budin (eds.), *Handbook of Terminology Management (Volume 2)*. Amsterdam & Philadelphia: John Benjamins.
- Ahmad, K. and M. Rogers, 2001. Corpus Linguistics and Terminology Extraction. In S.-E. Wright and G. Budin (eds.), *Handbook of Terminology Management (Volume 2)*. Amsterdam & Philadelphia: John Benjamins.
- Chang, S.-F. and J. R. Smith, 1997. Visual Information Retrieval from Large Distributed Online Repositories. *Communications of the ACM*, 40(12): 63-71.
- Flickner, M., H. Sawhney, W. Niblack, et al., 1995. Query by image and video content: The QBIC system. *IEEE Computer Magazine*, 28:23-32.
- Jørgensen, C., 1999. Access to Pictorial Material: a review of current research and future prospects. *Computers and the Humanities*, 33:293-318.
- Panofsky, E., 1939. *Meaning in the Visual Arts*. Harmondsworth: Penguin.
- Salway, A., 1999. *Video Annotation: the role of specialist text*. PhD Thesis, University of Surrey.
- Santini, S., 2001, *Exploratory Image Databases*. San Diego: Academic Press.
- Shatford, S., 1986. Analyzing the Subject of a Picture: a theoretical approach. *Cataloging and Classification Quarterly*, 6(3):39-62.

8. Acknowledgements

The authors are grateful for the suggestions of the anonymous reviewers and would also like to thank Lee Gillam who commented on an earlier draft. This work was partially funded by EPSRC through the SOCIS project - Scene of Crime Information System, (GR/M89041/01).