An open multilevel classification scheme for the visual layout of comics and graphic novels: definitions and use

John A. Bateman*, Francisco O.D. Veloso, Janina Wildfeuer*, Felix Hiu Laam Cheung** and Nancy Songdan Guo*****

*Department of Linguistics, Bremen University
**Department of English, Hong Kong Polytechnic University
***School of Professional Education and Executive Development, Hong Kong Polytechnic University

Version V1.2, June 2016

Abstract

Analytic interest in the functioning of comics, graphic novels and similarly visual media is currently experiencing considerable growth. In order to pursue empirical investigation of such media, it is useful to consider how data of this kind can be made accessible for the application of established empirical methods, such as corpus analysis. Up until now, the description of the particularly visual contribution of page layout and composition within comics has rarely received systematic attention, despite being noted by several authors to be an essential facet of the comics and graphic novels medium. In this paper, we motivate and describe a detailed annotation scheme for making this aspect of comics and graphic novels accessible to corpus methods. The paper sets out the general purpose of this endeavour, briefly summarises the state of the art, and then goes on to define the categories of our annotation scheme and illustrate its use for annotation.

Revision History

2015-12-19	Discussion of overfilling and underfilling refined
2016-02-04	Merged with coding instruction documents and updated to reflect the latest classification network
2016-04-20	Overlaps with DSH paper removed/reduced (V1.1)
2016-06-09	Further clean-ups of feature names for internal consistency
2016-06-11	Further examples and text clarifications

Contents

1 Purpose

When looking at comics and other forms of visual narrative, it is clear that there are considerable differences among them – but what are those differences? Can we characterise them and explore how they vary across time, geographical areas, cultures, artists, publishing houses and so on? In particular, there have up until now been few ways of objectively comparing the *visual appearance* of comics, even though this is the aspect of visual narrative that probably strikes the reader first most prominently. The current document consequently sets out a classification scheme for describing the *page layout* of comics. It is intended that any comics page should receive a classification according to the given classification network. Then comics from different time periods, cultures, artists and so on can be systematically compared to see if they use layout similarly or differently and in what respects.

2 General Background

Annotation schemes for linearly organised communication – such as, on the one hand, spoken language and dialogic interaction and, on the other, written verbal language – now constitute a well established tool for corpus-based research into communication. The situation with spatially-based media – such as comics and graphic novels – is very different (for further details, see Bateman 2014). Such media make extensive use of the two-dimensional space of the page in the creation of their media-specific effects and this requires rather different solutions to the challenges of annotation that have not been met sufficiently in proposals to date. Some of these proposals deal with comics and graphic novels by defocusing the visual contribution in favour of the verbal, as is the case in the Comic Book Markup Language (CBML: Walsh 2012), which extends notions from the well-established Text Encoding Initiative guidelines (TEI: Vanhoutte 2004) designed for computational support for text studies in the humanities; others provide only a geometric characterisation of page layout as often found in accounts building on, or feeding into, OCR technology, as in the case of the very useful corpus of annotated comics layout developed by Guérin et al. (2013). Our aim in this paper is to take a step beyond both kinds of approach by providing an annotation scheme that supports description more suited to investigations of narrative structure, aesthetic and perceptual impact, visual design strategies, and variations in visual style over time or across cultures, authors, publishers, etc.

In order to move the level of abstraction of description away from geometric information and towards distinctions that capture distinctive differences in the visual design options taken up, more qualitative levels of description are necessary. For example, the fact that two specific layouts include particularly sized rectangular areas demarcated within their respective pages does not of itself bring out distinctive aspects of a comic's organisation, such as the fact that particular grid organisations may or may not be being used, that some panels may be dependent on others or may even work as narratively relevant insertions or digressions, and so on. Generalisations are required over geometric data in order to separate significant distinctions from what may be, from the perspective of comics perception, less relevant metric variation. We suggest that such qualitative descriptions can be expected to make more direct connection with narrative and other high-level interpretations and so can be seen as constituting a further essential part of the bridge from physical realisation to the issues raised when addressing the communicative and aesthetic functioning of the comics medium.

Our goal in this paper, therefore, is to propose a detailed annotation scheme that addresses these concerns. The paper will motivate the scheme both in terms of the phenomena to be covered and in relation to existing schemes for describing comics and visual narrative. To begin, we will first briefly characterise some of the previous work in the area of providing annotation schemes for comics, clarifying which aspects of these schemes we draw upon and how extensions are necessary for qualitative visual descriptions. Second, we run through and briefly motivate the top-level organisational distinctions that our annotation scheme relies upon. And third, we illustrate the finer level distinctions of our scheme with respect to concrete examples. The paper then concludes with a brief illustration of using the annotation scheme for analysis and a summary of the next steps necessary to take this direction of research further.

3 Towards annotation schemes for comics

Work on comics and graphic novels from interpretative traditions such as narratology, literature and comics studies more broadly emphasises that the visual properties of the medium play a crucial role for its operation and effectiveness as a communicative form. However, characterisations of those visual properties remain by and large informal and subject to considerable variation. Details of visual organisation are typically mentioned only when prominent enough to warrant consideration in other more interpretative and discursive analyses. Such a basis is insufficient for broader, more corpus-oriented or data-driven characterisations of the comics and graphic novels media as a whole and the variations and developments found in these media across time and space.

It is also striking that the most common areas of discussion in static visual narrative research remain focused on panels – either in terms of their internal organisation or with respect to inter-panel relationships or 'one dimensional' structural configurations (Eisner 1992; McCloud 1994; Cohn 2010). Although certainly relevant both for analyses of comics in general and for annotation schemes for comics, such descriptions rely considerably on the linearity already established in accounts of language-dominated print media or verbal language. Several comics researchers do go further, however, and point out the significant contribution made by the use of the two-dimensional space of the page for stylistic, narrative and otherwise communicative effect as a specific property of the comics and graphic novel media (cf. Groensteen 2007 [1999]; Ludwig 2015). As Magnussen emphasises:

> "In reality the interpretation of the sequence is constant interaction with the visuality of the whole page. Even before beginning a sequential interpretation, an impression of the full page, or double page, has been made. A first overall scan is not merely confined to following for example a se-

> quence of actions but is free to be attracted by salient features like shape or colour, or by the content of relatively large panels." (Magnussen 2000: 200)

Systematic characterisations of this aspect of comics and graphic novels remain rare and, even when offered, are of limited generality (Kannenberg, Jr. 1993; Peeters 2007 [1998]; Caldwell 2012; Cohn 2013b; Horstkotte 2013). Developing an annotation scheme for this facet of comics and graphic novels is therefore a necessary task at this time. To provide such a scheme, however, we need to pay close attention to the actual forms that regularly occur in comics and graphic novels in order to bring out specifically those patterns which are held to be *distinctively* different from others. For this we will build in the present paper both on extant proposals for describing page layout in these media and on our own detailed explorative study of diverse types of comics and graphic novels from different periods.

For the technical realisation of this annotation scheme, our approach draws on established practice in other areas of (particularly linguistic) data annotation where considerable experience has been gained concerning methods for describing semiotically complex artefacts or behaviours. In particular, we divide the annotation tasks into several independent layers that are related both to one another and to the original documents by indirect, stand-off annotation. This allows us to construct differing perspectives on the artefacts analysed without giving one perspective or level of abstraction a determining role. This design decision for annotation is crucial both for doing justice to the complex interrelations observable in comics at all levels and for maintaining an open architecture which is always able to accommodate additional levels of description as these are found important for particular research questions addressed. For the purposes of the present paper, we will focus primarily on what we term the 'layout layer' of annotation, commenting as we proceed on some of the relations that are evidently required for capturing connections with other layers.

3.1 Previous annotation schemes

To begin, we need first to briefly set out previous proposals for comics annotations and some of the problems raised. We will build directly on the experiences of these approaches when attempting to provide more direct ways of representing comicbooks as data that can be investigated empirically in a manner similar to that used in other quantitative, corpus-based studies. As suggested in the introduction, previous annotation schemes for comics either address geometric aspects or page layout or have been oriented towards the textual component of comics.

The former, geometric approaches are relatively straightforward to describe in that they draw predominantly on rectangular or polygon areas and their coordinates within the page. Such information is useful for several purposes, perhaps most common however is its support of electronic distribution or of electronic editions, both of which are currently expanding rapidly along with the use of comics-focused eReaders; other uses of such geometric schemes are as sources of 'ground truth', i.e., data characterisations that may be taken to be 'true' for subsequent research purposes, for developing

```
<page title="Chapter 1: How it All Began">
    <image href="page1.jpg"/>
    <text-layer lang="en">
      <text-area points="10,10 50,10 10,100 50,100">
        <p> ... </p>
      </text-area>
    </text-layer>
    <frame points="10,75 650,137 650,562 10,562">
     ...
    </frame>
</page>
```

Figure 1: An example fragment of an annotation for a comic expressed in the 'Advanced Comic Book Format' (ACBF)

and evaluating automatic document recognition techniques. For uses of these kinds it is necessary to record specific data about the document including exact metrics about frame size, positioning, etc. Good representatives of such schemes are the Advanced Comic Book Format (ACBF: `https://launchpad.net/acbf`) and the form of annotation employed in the eBDtheque corpus of comicbooks layouts (Guérin et al. 2013) under development by the Informatique, Image, Interaction group at the University of La Rochelle (`http://ebdtheque.univ-lr.fr`).

As we shall explain in more detail below, annotating predominantly visual material where the original data is commonly represented as a graphic image is best approached using 'multi-layer' annotation schemes. ACBF, for example, separates out the base image level, which is a simple graphic file, a text level, in which the text of captions, speech balloons and so on is recorded, and frame information, that sets out the geometric positioning of panels. Information in the scheme as a whole is expressed using the extensible markup language (XML), as will be the case with all the examples we discuss. A corresponding example adapted from the ACBF specification is shown in Figure 1. Here we can see that the original comics page is picked out by a reference attribute in the `image` tag (line 2) and areas within that image are picked out by series of coordinates, either within the `text-area` tag for textual information (including a usual range of typographical formatting options) or within the `frame` tag for visual subareas. The frame elements are used during reading or browsing in order to define regions that may sensibly be zoomed, such as panels.

In many respects, this is the usual scheme offered in such annotation standards as the Text Encoding Initiative (TEI: see below) for dealing with works where the visual layout is more complex and requires decomposition into areas of interest of various kinds. Separating levels or information in this way makes it easier, for example, to search for particular textual phenomena, and so supports indexing and retrieval, as well as the possibility of incorporating translations for multilingual comics versions. Such page specifications are themselves embedded among further document annotations, such as metadata concerning authorship, publication details, genre, characters

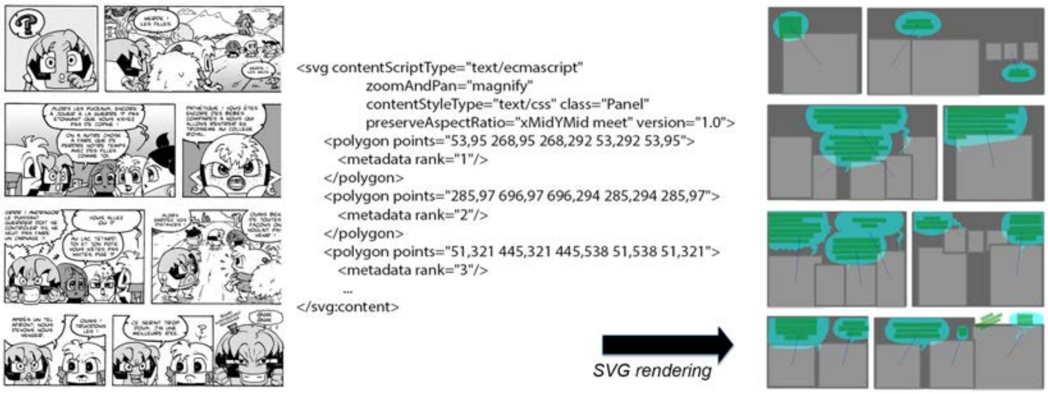


Figure 2: eBDtheque annotation of a comics page (left) in terms of scalable vector graphics (middle) supporting segmentation visualisation (right). The page is from 'La Légende des Yaouanks' from Cyborg 07 (2009), 'CYB_BUBBLEGOM_T01_005' in the eBDtheque corpus.

that appear, and so on. In the particular case of ACBF, the purpose is to provide enriched structural and semantic information concerning the comics represented and to show reading orders that may be employed by comicbooks readers for display.

Whereas ACBF is aimed for electronically distributing comics and for organising comics collections, a rather different set of purposes is pursued in the eBDtheque scheme. Here the primary task is that of providing reliable data for the evaluation of automatic layout recognition systems. The eBDtheque scheme is also expressed in XML but in this case relying in particular on the Scalable Vector Graphics (SVG) format. This then also directly supports visualisation of the annotated areas using standard SVG rendering tools. An example from the database (Version 2, June 2014) is shown in Figure 2. Similarly to the ACBF scheme, this format specifies the annotated page as an image and defines variously shaped polygons to be placed 'on top' of this image. In this case, particular styles for visualising the specified polygons are given by cascaded style sheet (CSS) information allowing ready modification of the visualisation (as, in fact, we have done in the figure in order to emphasise the panel segmentation). Again, we see here that the definition of panels as polygons is the primary annotation information provided that is related specifically to comicbook layout, although eBDtheque now also goes further and provides very detailed renditions of the shapes of speech and thought balloons as well as subregions within panels annotated to identify the story characters that appear in the comic (also visible in the panel internal rectangles shown in the figure).

Some of the features of these annotation schemes are quite general and can also be adopted in our own scheme – metadata organisations for comics, for example, are already appropriately defined and so can be adopted without change. The reliance for access to the visual material on reduction to geometric regions is, in contrast, rather too limited for purposes of further analysis. A very different line of annotation development is accordingly taken by the 'Comic Book Markup Language' scheme (CBML:

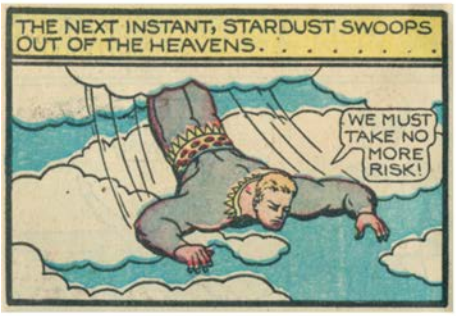
Walsh 2012), which as remarked above follows established and standardised annotation practices from the Text Encoding Initiative framework (TEI: Vanhoutte 2004). One goal of the TEI is to provide structured representations for a broad range of documents capturing the functional and structural organisation of those documents and making this organisation computationally accessible. The application of this strategy is therefore an important step away from more straightforward geometric annotations, providing more explicit characterisations of comics-specific organisations.

In practice, however, the particularly comicbook related extensions made in CBML remain rather few and most of its concerns and methods are inherited from the TEI. The most important comics-specific extensions are the following, selected with respect to the CBML version formally defined at `http://dcl.slis.indiana.edu/cbml/schema/cbml.odd`):

- <panel> demarcates a single comicbook panel. The tag defines attributes of characters, listing the characters in the panel with identifiers, a panel number and a panel type. The tag also inherits from the TEI specification the global attribute type 'ana' for associating interpretations with elements. Walsh gives an example usage of this attribute for recording panel relations of the kind set out by McCloud (1994), i.e., moment-to-moment, action-to-action, subject-to-subject, scene-to-scene, aspect-to-aspect and non-sequitur.
- <balloon> characterises a single speech, thought, or other type of balloon, as commonly found in comics, comicbooks, and graphic novels. The tag defines an attribute type according to whether speech, thought, audio, or telepathy is being used, and further identifies the source (sayer, thinker, etc.) from whom or which the balloon content is originating.
- <caption> picks out a narrative caption from a comicbook panel, inheriting the attribute 'rendition' for capturing basic typography (such as upper case, etc.).

Use is also made of some of the standard TEI-defined tags by applying these to comics-specific cases. Examples of these include:

- <ab> (anonymous block) contains any arbitrary component-level unit of text, acting as an anonymous container for phrase or inter level elements analogous to, but without the semantic baggage of, a paragraph (taken from the TEI module 'linking')
- <div> (text division) contains a subdivision of the front, body, or back of a text (taken from the TEI module 'textstructure'): extended for CBML when the type is 'panelgrp'. This is generally used to represent collections of panels on a page; as Walsh describes it: "Panels are intentionally grouped and composed in larger compositional units, often corresponding to a physical page surface." This tag is also, however, used with other types in order to capture other divisions within the document, such as chapters and so on.
- <floatingText> is used to capture text that is present within the panel, such as might be displayed on posters or signposts or other diegetically relevant objects depicted.



A panel from a Stardust story, by Fletcher Hanks, from Fantastic Comics #15 (February 1941).

```
<cbml:panel
    characters="#stardust"  xmlns:cbml=
    "http://www.cbml.org/ns/1.0"
    <cbml:caption>
        The next instant, Stardust
    swoops out of the
    heavens........
    </cbml:caption>
    <cbml:balloon
        type="speech"
        who="#stardust">
        We must take no more risk!
    </cbml:balloon>
</cbml:panel>
```

Figure 3: Example of panel annotation employing CBML

An example of CBML markup for a comics panel offered by Walsh is given in Figure 3. The annotation identifies the graphic resources employed in the panel — here a caption and a speech balloon — as well as the comics characters depicted in the panel and who may be responsible for any speech or thought depicted. This is indicated by means of designated values of the 'characters' and 'who' attributes respectively. All of this information can then subsequently be searched or indexed – as, for example, when we want to examine all panels containing a particular character or to examine the language associated with any character. The annotations provided also support further exploration of the features of the medium, such as the distribution of captions, speech and thought ballons and so on, over time or across authors, countries, publishers, etc.

There are, then, clearly many ways in which explicit representation of this kind in order to capture information about a comic or collection of comics can be beneficial for subsequent research. However, there are also drawbacks inherited from the rather straightforward adoption of TEI practices. In particular, although the use of tags defined from those that TEI offers conforms to the TEI intention that a generic scheme for documents of all kinds is provided, less clear is the extent to which the comics-specific nature of the phenomena to be addressed here is effectively covered. Indeed, in some cases, the import of TEI categories and their definitions – and, in particular, the interdependencies that the now rather complex TEI definition creates – turns out to be somewhat counter-intuitive for addressing the particular concerns of the comicbook and graphic novel media.

One example of this can be seen in the characterisation of 'sounds' depicted in comics panels. For this CBML employs the standard TEI tag:

- <sound> describes a sound effect or musical sequence specified within a screen play or radio script.

This tag is taken from the TEI module 'drama' where sounds may need to be specified as stage instructions. In order to cover the full range of such instructions in documents in general, the TEI specification defines this tag recursively to allow sound elements within sound elements. Although no doubt necessary for certain documents, for comics the utility of this modelling decision is less clear – what sounds within sounds may mean for a comicbook is difficult to see (hear). Moreover, sound elements may also legally (i.e., following the usage CBML makes of the TEI definitions) contain balloons and captions. This degree of interdependence and recursivity is of questionable applicability for the purposes of comicbook description. The difference is that the elements that express sounds in comics have very specific characteristics which are not then necessarily compatible with uses in other media (cf., e.g., Cohn 2013a). Indeed, <sound> in the context of drama may *describe* sounds and so may be arbitrarily complex; in comics the associated elements do not describe sounds, they *are* sounds (or are to be read as such – as in BANG!, CRAAAAASH!, and so on). This is quite a restricted (and comics-specific) function.

The same problem applies to the *attributes* that may be used to specify additional information about some element. For example, the CBML tag 'panel' also formally allows a considerable range of TEI elements that are not specifically related to comics and whose applicability is questionable. Many of the constraints imposed by the TEI are of this form, which detracts from its potential to be usable not only for markup but for reasoning about the knowledge structures thus created. What is required is a far tighter modelling of the particular combinations of values and elements, and their structures, that picks out just those distinctions that are relevant for the class of documents that includes comics and graphic novels.

In short, the TEI is well motivated for textual information, for predominantly textual documents and for extracting predominantly textual information from visually more sophisticated artefacts. For strongly visually organised documents, however, such as comics, where the spatial organisation of the artefact itself makes a substantial contribution, the approach is relatively rudimentary. This then leads to the CBML specification also being less than ideally suited for several central areas necessary for comicbook annotation. At the same time, many aspects of encodings required for text documents in general are less relevant for comics but nevertheless find their way into the CBML by virtue of import from the TEI specification, reducing the usability and precision of any resulting annotation.

3.2 Requirements for comics annotation: modularity and layers

When deciding on an overall annotation framework within which to couch our proposals for capturing comics and graphic novel layout, it is then relevant to consider to what extent current approaches drawing on TEI are appropriate. A particular concern we have raised is the inheritance of potentially inappropriate characterisations that violate medium-specific aspects of some class of documents. Indeed, within the standard document annotation models employed with TEI, all TEI components are 'passed on' to submodules. This means that there is actually little formal modularity in the resulting specifications. Constraints between tags consequently err on the side of allowing far more than is going to be found in any document and range freely across modules. Remaining within the standard TEI framework for annotating comics and graphics novels may not then be an ideal choice.

This conclusion is not, moreover, specific to comics and similar moves have now been made in several areas. In particular, work on richly annotating *linguistic* data has shown that best design practice involves a 'multi-layer' approach to annotation, where data is represented from several perspectives, or layers, and each such layer is subject to its own definitional requirements and is responsible for particular kinds of information. There is no reason, for example, that annotations of intonational information need to respect the tags and properties defined for annotations of syntax or of typography. Several annotation frameworks and even tools supporting such annotation architectures for complex linguistic data are now available (cf. Dipper 2005) .

Problems of this kind consequently reflect a more general challenge for information modelling – that of being able to define and use appropriate specifications of *modularity* in that modelling. A lack of modularity is problematic for the separation of ontologically distinct kinds of information: for example, many of the drama tags may be relevant to describe the content of the unfolding storyline within a comic book, but the separation of this level of description from more realisational tags, such as the 'sound' tag mentioned above, makes the resulting annotation overly complex. There is also the converse problem of over-specificity: particular types of units are suggested but these are not clearly demarcated as the main places where theoretical distinctions are to be argued and debated: the example mentioned above of employing McCloud's inter-panel relations is one case of this – the kinds of relations that might be found between panels is a major component of any account of the workings of comics, but the annotation relegates this to an unstructured list of possible fillers for an attribute that must still then be specified as part of the annotation scheme.

A similar example is the provision made for a certain number of balloon types, also maintained as a more or less arbitrary list that is nevertheless part of the annotation scheme. Placing such lists in an annotation framework is often not a good choice since they may well be constructed textually within single documents – the 'telepathy' class mentioned above shows this well since how precisely a telepathy 'balloon' is formed may be document-specific. More useful for a framework would be to allow classes to be constructed dynamically within the interpretation of even single works rather than incorporating these as extra baggage for analysis frameworks as a whole.

Walsh notes this flexibility and gives examples of how conventions for the representational styles of particular balloon types can emerge and evolve. This is not sufficiently represented in the annotation scheme, however. Such information is only captured within the 'rendition' tag, which is intended to contain traditional typographic information:

> "In primarily textual documents, the <rendition> element and related attributes might be used to describe details such as font family, font size, justification, and so on. The rendition features of TEI may be used in CBML contexts to describe graphical features, for instance to provide a detailed description of the distinctive styling of the android's speech balloon." (Walsh 2012: 34)

How this visual aspect is to be covered is not addressed in any detail; examples suggest that such descriptions may be provided in natural language glosses commenting particular items identified in the annotation – allowing open commenting, however, is clearly only a small step towards the formalisation necessary for annotating comics and graphic novels reliably and a more precise coding practice would be preferable.

Attempting to combine such heterogeneous kinds of information in a single annotation scheme therefore tends to lead both to over-complex maintenance and representation mechanisms and to less well defined annotation schemes. This problem has also been addressed in current work on empirical multimodal corpus-based research, where much of value for considering comics annotation can be found. A more productive approach will therefore be to combine the insights of both, particularly with respect to the accumulation of multiple layers of annotation. Thus, although all of the schemes described so far offer benefits for dealing with the medium of comics using corpus-based and other automatic processing methods, they are also inherently restricted. Moreover, although such schemes may also be employed for research and study of how comicbooks work as a medium, this aim also brings it own constraints to bear and it is useful to consider whether tools tailored more specifically to the goal might be defined.

We should also note that, although stand-off annotation is clearly a prerequisite for the kind of annotation scheme we require, the TEI has in general had a difficult relationship with such accounts (cf., e.g., Bánski 2010). Despite it now being broadly accepted that such annotation techniques are essential for complex annotations for much of the data falling within the remit of TEI, there is still no established component of the standard for expressing sets of annotation relying on stand-off techniques. This weakness goes back to the 'ordered hierarchy of content objects' (OHCO) assumption that informed much early thinking on annotation (cf. DeRose et al. 1990). Although sufficient for many purposes, when considering annotations motivated by more complex theoretical models it is common that there are multiple hierarchies of differing kinds of descriptions and that those hierarchies need not align, even defining differing and possibly overlapping structural units. This in turn violates many of the simpler 'import' models for combining specification modules.

The OHCO assumption means that stand-off annotation can be seen as a kind of 'shorthand' for more detailed descriptions of a document – the stand-off annotation can under this view always be 'folded back in' to the source document giving a fully articulated version without stand-off annotation. As soon as the OHCO assumption is removed, this model of import no longer functions. 'Folding in' stand-off annotations cannot be expected to result in a single well-formed XML document because of the possibility (indeed, the likelihood) that cross-cutting units will be required. For linguistic annotation, this property is already seen as essential – for mixed semiotic artefacts, the challenge becomes even greater. Annotation needs to be seen as levels of descriptions rather than as ways of characterising structured 'content objects'.

The provision of techniques and mechanisms supporting these more complex annotation and import methods is an active area of debate within the TEI community and no doubt useful guidelines will emerge (cf. Pose et al. 2014). For our current purposes, however, we already need to rely on multi-layer annotation schemes and so must commit to some particular annotation forms. Therefore we assume a relatively simple multilayer organisation consisting of the raw data to be annotated as a graphics image, a separate labelling layer, called the base layer (see below), providing identifiers for areas within the image, and an open set of further layers providing information and structure over and above the labelled elements of the base layer. Stand-off annotation is realised by always referring indirectly to areas within the image as regions or to already labelled base units within the labelling layer. This annotation scheme is represented throughout in XML so that it may be convertible to any more specific stand-off annotation scheme, including emerging standards, as required. The details of the open layer stand-off framework that we assume builds directly on our previous experience in designing annotation schemes for the visual aspects of other kinds of printed documents, particularly as pursued within the Genre and Multimodality (GeM) project (Bateman 2008) – we return to this in Section 4.

3.3 Towards a multilayer scheme for the visual organisation of comics

We have suggested that an orientation towards TEI textual annotation without stand-off layers leads in the case of CBML to several unfortunate consequences for the treatment of particular types of documents when those documents differ substantially from the text-oriented prototypes foreseen for TEI's application. In particular, CBML's foundation on TEI has led to its inheriting a strong *textual* orientation. In certain respects, this is certainly appropriate – since comics books are published as books and share many of the physical features and logical properties of production with books, the corresponding TEI elements are without doubt relevant. But for comicbooks we are also crucially involved with investigating relations among *visual* media and this entire facet of comicbook design is necessarily de-centred in a TEI-based representation such as CBML.

Walsh (2012) suggests this is justified, and has even been argued in work on comics (cf., e.g., Abbott 1986), by the fact that in everyday usage one talks of 'reading a comic'. But to go from the notion of 'reading' to an assumption of the applicability of a text-oriented annotation approach is inappropriate for at least two reasons: first, since the *medium* of comics has until recently been traditional print, reading is uncritically accepted as just the kind of activity that goes with 'books' – even the probably contradictory phrasing of 'I read a picture book' is unlikely to raise eyebrows. This does not mean, however, that one is necessarily approaching pictures (or combinations of pictures and texts) in the same way that one is approaching simple text documents and so a more critical stance needs to be taken. Second, the use of 'reading' is itself a consequence of the cultural prominence given to text since the late nineteenth century (Kress and van Leeuwen 2001: 1) and similarly says very little about the artefacts being examined – we simply lack a corresponding verb that covers detailed visual appreciation/interpretation. In both cases, an appeal to the use of 'reading' to decentre the visual component of comics is unconvincing.

This means that it is by no means possible to subscribe to the following restriction suggested by Walsh for annotation:

> "...the examples above do not attempt to describe the pictures one finds in the comic book, nor should they. Comic books are a visual, graphic art form combining text and image.

> CBML/TEI/XML is a text format. While one could certainly use CBML to describe details about any or all of the pictures in a comic book publication, such an effort would undermine the hybrid form of the comic book. The visual, pictorial, and graphic design elements of the comic book simply cannot be fully or adequately described or translated as text." (Walsh 2012: 48)

This conflates the use of annotation with the subject matter which is being annotated – again because of the original purpose of TEI of adding information to textual documents. Nowadays in empirical multimodal research it is beneficial to provide markup for all kinds of information, not just textual. Thus while "CBML/TEI" may be a text format, "XML" certainly is not: it is a scheme for adding information to *any* kind to data. And this is precisely what we need to do in order to pursue empirical investigations of multimodal artefacts, such as comics.

Walsh then continues:

> "it would be futile and impractical to attempt to describe every detail of every picture in a comic book document." (Walsh 2012: 48)

This is certainly true. The goal of annotation is not, however, to "describe every detail" but to achieve useful levels of analytic abstraction that provide information *about* data, not to *reproduce* that data. This conflation of goals can be traced back to earlier uses of the notion of 'transcription' where, due to technical limitations, researchers often worked with transcribed data because the original data was too difficult to gain access to on its own (cf. Ochs 1979). In the visual case, there is no need to describe the picture when we can simply reproduce it; there *is* a need to analyse the picture, however – otherwise we cannot explore how it is that meaning is being made.

Walsh notes that such information can be added into TEI and, hence, CBML documents using extra attributes such as 'note', whereby free text comments can be associated with elements, and 'ana', which as mentioned above is used to explicitly link to analyses and interpretations. This moves in the right direction but still sees analysis as a subordinate goal of the markup. This is appropriate for providing electronically enriched editions of documents, since there analysis is certainly not in the centre of attention, but it is not appropriate for supporting analyses in their own right.

This change in the function of annotation is explained and illustrated in more detail for empirical multimodal analysis in general in Bateman (2013). Indeed, as Walsh himself concludes:

> "I have argued above that CBML should not attempt to go too far in description of individual images and pictorial details, relying instead on the presence of facsimile page images. Nonetheless, in order to analyze the visual grammar and conventions of comics, additional visual and pictorial features will need to be "identified explicitly in order to facilitate processing by computer programs" [TEI 2010c]. Future work on CBML will include modeling frameworks for analysis of pictorial and graphic features and developing taxonomies for identifying such features, beyond the basic structural components of panels, balloons, captions, and sound effects." (Walsh 2012: 62)

The question remains as to how best such visual extensions might be included, which is the task we take up directly in this paper.

4 Multilayering and the GeM model

The need to provide information about non-textual aspects of artefacts that is identified by Walsh occurs far more broadly than in considerations of comicbooks. Nevertheless, this more visual perspective has still received insufficient attention within annotation schemes in the TEI tradition. This is understandable since the development of TEI originated in work on linguistic corpus-based analysis and digital editions for the humanities, all heavily text-dominant forms. In contrast, most modern print-based and online documents exhibit an equally strong, if not stronger, visual orientation and several research projects have been considering how the needs of capturing documents of this kind may be met.

One approach aiming to deal with this situation and to extend corpus-based empirical methods to non-textually dominated documents is the GeM markup scheme developed within the Genre and Multimodality project (cf. Bateman et al. 2004; Bateman 2008). The goal of this project was to explore a characterisation of document types by extending the linguistic notion of genre across media and including visual document styles. Consequently, the approach developed in this project takes the visual organisation of multimodal data, specifically multimodal documents, as central. From semiotic, reception and production perspectives this is an appropriate path to follow because the visual is the most immediate mode of access of readers to such mixed text-image artefacts and is, for many artefacts, a principal consideration during design.

This orientation strongly prioritises the 'page' as it is physically manifested as a semiotic resource for communication in its own right. In this respect, Walsh's (2012) following citation of Rust (2008) is particularly relevant:

> "Both medieval book artists and contemporary cartoonists make use of the page as a device for giving their readers access to a domain of representation that is beyond the regimes of either pictures or words – yet somehow in the shadow of both." (Rust 2008: 25)

The GeM approach was specifically aimed at bringing the contribution of the page into the forefront, allowing both text and visuals equal status within the annotation scheme and, moreover, to show how these modes can interact in the service of providing more meaning than is present in each alone. This is then a very different starting point to the framework supported by the TEI, which succeeds in placing both visuals and the page as such entirely in the shadow of the text. The difference in orientations is suggested graphically in Figure 4, which shows the relationship between three organisational modes for page information explored empirically within the GeM project — text-flow, image-flow and page-flow — and the kinds of documents covered by the TEI. Document and page organisations that are strongly oriented towards manipulating the two-dimensional space of the page as a semiotic resource are covered under 'page-flow', which lie outside the capabilities supported by the TEI to date.

It is, furthermore, common for visually-oriented artefacts to involve many layers of organisation which cannot be readily annotated within a single level of description as the TEI scheme would favour. For this reason, the GeM model is an open multilayered annotation scheme more related to current approaches to annotating complex multimodal data of other kinds, such as face-to-face interactive dialogue. The multilayered

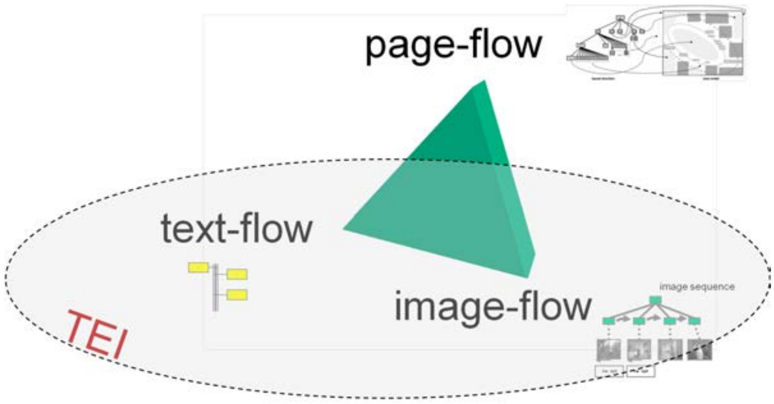


Figure 4: Relation between the kinds of document organisation addressed by the GeM project and those covered by the TEI

approach thus includes distinct annotations for a variety of kinds of analysis. In the case of documents, this allows visually-based annotations to exist alongside textual annotations, and both to exist alongside more abstract text and document organisational levels of description. This is the feature of the scheme that we will employ particularly below. It would also make it possible, for example, to import the entire CBML contribution as one level of more text-oriented information for any comicbooks analysed, thus allowing one of the primary limitations of CBML to be addressed while still drawing on some of its proposals.

The GeM scheme was not developed with comicbooks in mind (cf., e.g., Thomas 2009; Hiippala 2012); in this paper, therefore, we focus on the extensions which we are currently designing in order to provide a robust empirical foundation for more visually-aware investigations of comicbooks as well. We present this largely independently of other layers so that reuse and combination with other annotation schemes, including TEI when an agreed treatment of stand-off annotation has been established, can be considered.

The core levels defined by the GeM model relevant to our discussion here are:

(i) a base level assigning identifying labels to all elements on a 'page' that are to receive annotation – this is the basis for the multilayer 'stand-off' annotation employed; all subsequent markup proceeds via these identifiers rather than in the original document itself (cf. Bateman 2013);

(ii) a layout level that captures the *visual* segmentation of a single 'page' or 'spread' – i.e., that unit that a reader typically encounters as a whole within the visual field;

(iii) a rhetorical level that describes the intende communicative relations between distinct layout elements within a page;

(iv) a navigation level for capturing page cross-references and hyperlinks.

We will take the base level more or less for granted as it is straightforward both in conception and use; other levels not of concern presently include linguistic levels containing detailed syntactic and semantic analyses. The layout level is then the level that particularly requires attention and extension for application to comicbooks, while the rhetorical level may need adjustment to other, more comics-specific requirements such as, for example, kinds of discourse relations illustrated for comics by Bateman and Wildfeuer (2014) or the 'visual narrative' categories proposed by Cohn (2013d).

The documents investigated using the GeM scheme formerly motivated a characterisation of the layout level made up of three main kinds of information:

(ii.a) a layout structure, giving a hierarchical decomposition of a page organised around visual containment, visual proximity and functional relatedness;

(ii.b) a page model, giving a spatial representation of the page in terms of more or less complex page grids;

(ii.c) a realization level, showing the typographical properties of the various units of the layout structure.

A complete description of a page then consists of a layout structure whose elements are linked both to realization units and to particular spatial positions defined within the associated page model. Of these, clearly (ii.c) overlaps with the specifications offered by CBML and so need not be duplicated; in fact, the original GeM model also simply imports typographic markup from other schemes at this point (in particular, those of XML cascading style sheets and formatting objects). Within the representations developed for the documents originally addressed within the GeM framework, the realisational information assumed for layout units within the layout structure therefore commonly included specifications of the typefaces used within a unit, the type of unit (e.g., as text or image), colouring or framing and so on. These aspects do not then call for new additions at this stage.

Applying this framework to comicbooks for carrying out a basic layout annotation within the existing GeM framework then gives descriptions of the kind suggested graphically in Figure 5. The basic method in such annotations, illustrated in depth in Bateman (2008), is to perform a visual decomposition of the units found on a page and to characterise the result as a hierarchical structure of 'layout units'. This structure attempts to preserve the visual dependency relationships observed in the page, explicitly representing containment and other kinds of dependencies in terms of structural constituency within a layout structure. The layout structure may also include relative measurements or proportions at each level in the hierarchy — thus, for example, the first row on the left-hand side includes the information "50:50" for the spatial allocation of its children, whereas the second row includes the information "33:33:33". Although certain aspects of this layout characterisation can be carried over to comicbook layout, it is clearly not yet sufficient on its own and will need to be refined further below.

The task of producing visual layout structures of the kind shown here also overlaps in many respects with the standard task within automatic document recognition and classification of decomposing scanned or otherwise produced low-level representations of visual documents into hierarchical document structures (cf., e.g., Namboodiri

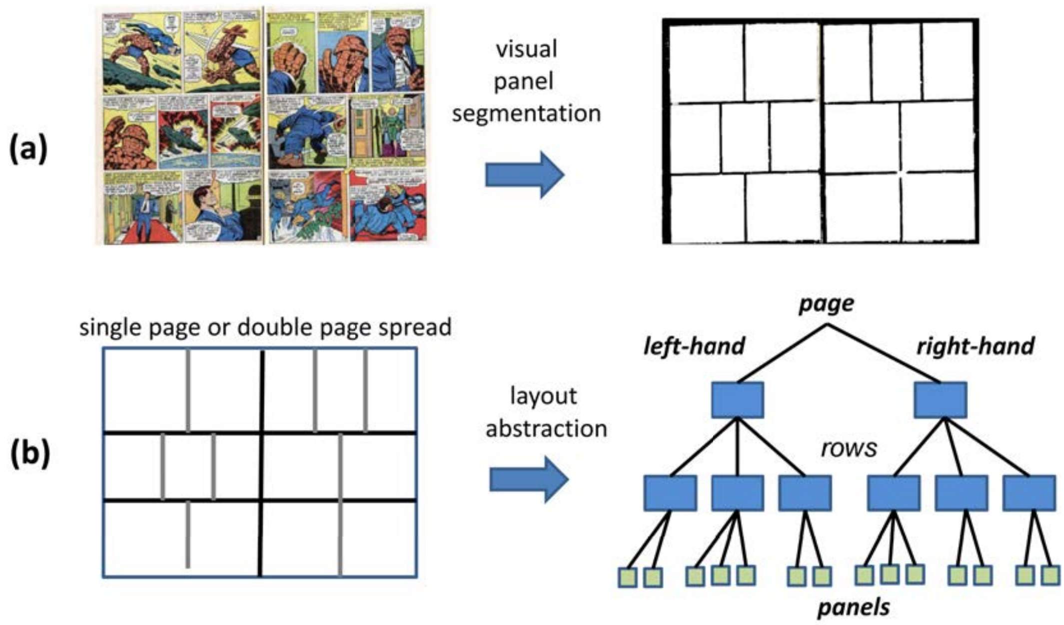


Figure 5: First two stages of annotation using the GeM layout level of annotation

and Jain 2007); this then makes a further connection to the comicbook annotations in projects such as eBDtheque introduced above.

Several techniques have been proposed for this process, ranging over applications of Voronoi representations for finding maximally connected 'areas' in the low-level representations (Kise et al. 1998), texture and edge differentiation models (Sumathi and Priya 2013), attention models (Maderlechner et al. 2000), and tessellations of various kinds (Xiao and Yan 2003). For the purposes of the GeM framework it was necessary to address how the results of such visual recognition processes could be best represented as part of the recorded visual layout structure.

Many approaches employ for this the notion of XY-trees, first introduced for automatic document processing by Nagy and Seth (1984). XY-trees operate by successively segmenting a visual layout by finding maximal partitions of a layout that successively switch between horizontal and vertical dimensions — that is, first a maximal partition is sought along the horizontal dimension and, subsequently, maximal partitions are sought for the segmented areas along their vertical dimensions, and so on recursively.

The result of this process is a visually-driven decomposition of the units found on a page that characterizes layout in terms of a hierarchical structure of 'layout units'. The operation of this technique for an example comics page is shown in Figure 6. A variety of further developments of XY-trees have now been explored (e.g., Cesarini et al. 1999, 2001; Sutheebanjard and Premchaiswadi 2010) as well as a range of further approaches sometimes used instead of XY-trees, sometimes as complementary approaches. A useful listing of several of these can be found in Rigaud et al. (2015: 202–203).

The layout structure shown for our example comicbook page lower-right in the Figure 5 corresponds to the XY-tree for the page. Moreover, very similarly moti-

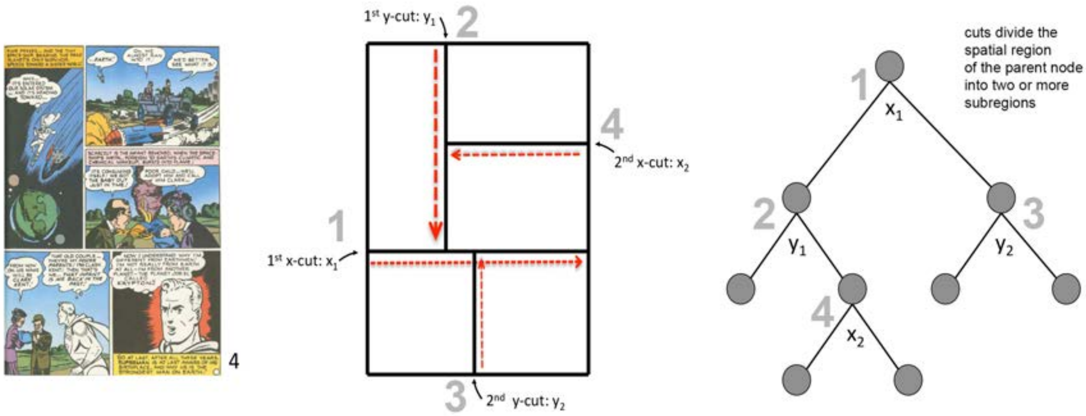


Figure 6: Example of construction of an XY-tree by successive horizontal (x) and vertical (y) ‘cuts’ of the original page layout. Each cut (x_1, y_1, x_2, y_2) divides a spatial region into subregions respecting the whitespace created by gutters, margins, etc.

vated structures have been suggested independently for comics by, for example, Cohn (2013b,c) and Chavanne (2015) as discussed in more detail in Bateman et al. (2016). Enforcing from the outset more structure on the page annotation in the way proposed in the GeM framework and as will be taken further below is quite valuable. It becomes possible via the associated realization level to capture commonalities in design features that may hold for entire groupings of panels. This has been suggested by Groensteen (2007 [1999]) and others to be fundamental to the medium. In addition, it is equally possible to explicitly locate patterns. For example, returning to Figure 5, we can readily describe similarities in visual structure and segmentation across the first and third rows on the left-hand page and the last two rows on the right-hand page: all have only two ‘children’ in their respective trees. Such similarities are straightforward to find with current query mechanisms and offer more information than simple numbering. For reference, however, we may still want to refer to individual panels; this is managed with the following structured numbering scheme:

$$\text{page \# . } \{\text{L, R, D}\} \text{ . row \# . panel \#}$$

For the example in the figure and for relatively simple grid organisations of this kind, we would then refer to the panel in the middle of the page on the left as panel: “18-19.D.l.2.2”. The XY-scheme allows generalisations beyond rows and columns, however. The panel midway down on the righthand side in Figure 6, for example, would be “4.L.1.2.2”, counting with respect to the XY-tree shown to the right of the figure – i.e., page number 4 (taken from our corpus numbering for expository purposes), which is a lefthand page, and within this the second branch of the second branch of the first branch of the XY-tree. Representations will generally be maintained for entire double pages since it is these units that may well be viewed by a reader in one ‘take’ and so there may be visual effects holding for the spread as a whole, even when the individual pages are intended to be read in sequence.

Finally, it is also worth emphasising at this point that it is *not* the case that the ordering of sibling elements below a node in the GeM layout structure corresponds to the order of ways in which those visual units are attended to by readers or viewers. That is, reading order is *not covered and must be additionally specified explicitly for any node*. This is evidently of particular importance for the more complex layout structures required for comics that we describe below and relates also to Cohn's discussion of how variations in layout structures may correlate with particular reading paths followed by readers (Cohn 2013b,c).

While more or less straightforward grids can be captured in the existing scheme, however, the resulting representations still fail to bring out particular layout qualities that are intrinsic to the comicbook medium and are also overly limited with respect to the variations in design that comicbooks regularly employ (and, moreover, have employed for a considerable portion of their history). To cover such possibilities, we will need to make substantial extensions to the descriptive possibilities of the layout structure. Even traditional simple panel layouts found in comicbooks can diverge from basic grid structures and so require refinements to the descriptive apparatus. Consider, for example, the pages shown in Figure 16 below: these are all far from simple grids, although clearly related in form. And, particularly in more recent comics, there are considerably more challenging layouts — as Walsh notes:

> "Simplistic approaches to encoding the spatial and sequential relationships among such panels are foiled by frequent variations and complexities in panel size, shape, and arrangement and particularly by ambiguous sequential positioning of panels." (Walsh 2012: 53)

For these reasons we now present a refinement and significant extension of the layout layer of the GeM annotation scheme specifically tuned to the needs of capturing the complexity and variation of comicbook visual organisations.

5 A visual layout scheme for comics: organisation and analysis units

In this section we undertake our primary task, setting out our proposed classification scheme for the visual appearance of comic book page compositions. To follow the motivations offered for the categories being distinguished, it is important to recall that we are focusing *solely* on the overall page composition and its organisation of panels, groups of panels and other visual material that might occur to bind these panels and groups together into compositional wholes. As suggested above, this is a facet of comics and graphic novel organisation that has received insufficient attention to date, even though it demonstrably impacts on interpretation, reading strategies and effect. Other essential contributions to artefacts of this kind, such as the content of individual panels and the visual styles exhibited by those panels, are to be captured by other levels of classification and annotation. The provision of an additional level of classification specifically for page layout is thus intended to support more effective study precisely of the *interdependencies* among such design choices and other components of the medium.

The method used to develop our classification scheme draws directly on the approach long established for language descriptions within systemic-functional linguistics (cf. Halliday and Matthiessen 2004). As a functional linguistic theory, systemic-functional linguistics focuses attention on language as a *resource* for meaning-making. This resource is characterised in terms of classificatory 'networks of choice' with associated structural consequences, called realisations. This form of deeply nested classification differs somewhat from many more traditional annotation schemes where the categories that are applied to any unit of analysis are generally 'flatter' with few subcategories. For those unfamiliar with the more hierarchically organised forms of annotation/classification, therefore, the following subsection gives some more background illustrating the linguistic use of classification networks of this kind. We then continue with the application of this approach to annotating comics and visual narratives.

5.1 Classification networks

Within systemic-functional linguistics, classification schemes for describing any level of linguistic abstraction (grammar, semantics, etc.) consist of networks of choices. Choices are characterised as networks because there are many choices that lead on to further choices that need to be made. Moreover, some choices need to be made in parallel – i.e., they are all necessary rather than being mutually exclusive. A description of a unit of analysis – be that a clause, a nominal phrase or, as will be the case for us below, a comic book page or spread – is complete when all of the choices possible have been made. The result of an analysis is a list of the choices – called 'features' – that have been selected. [1] It is these lists of features that can then be compared across time, across publishers, across countries, etc.

Networks as used in systemic-functional linguistics are referred to as *system networks*. We use a brief example from grammar to show how this form of classification works. A system network is a graph consisting of choice points called 'systems'. A small section of a system network from the grammar of English is shown in Figure 7. This subnetwork consists of eight systems, whose names are capitalized. The square brackets in the network represent systems and their choices. For example, the PROCESS TYPE system has four choices: 'mental', 'verbal', 'relational' and 'material'. If one gets to this system (by making previous choices to the left of the system as given in the graph), then one needs to select *just one* of these four.

Any choice that is made may then lead on to further choices – for example, if we choose 'indicative' in the system MOOD TYPE then we need to go on and make a choice in the INDICATIVE TYPE system. Curly brackets – for example as leading on from the choice of 'relational' – mean that one has to make independent, or parallel, choices in all of the systems given (e.g., one choice from RELATIONAL TYPE *and*

[1] In more traditional linguistic terms, this may be summed up by stating that systemic-functional linguistics places its main focus on paradigmatic organisation, whereas most structural frameworks focus on syntagmatic organisation. Both paradigmatic and syntagmatic descriptions are essential components of any complete account: the difference is one of organizational and methodological focus.

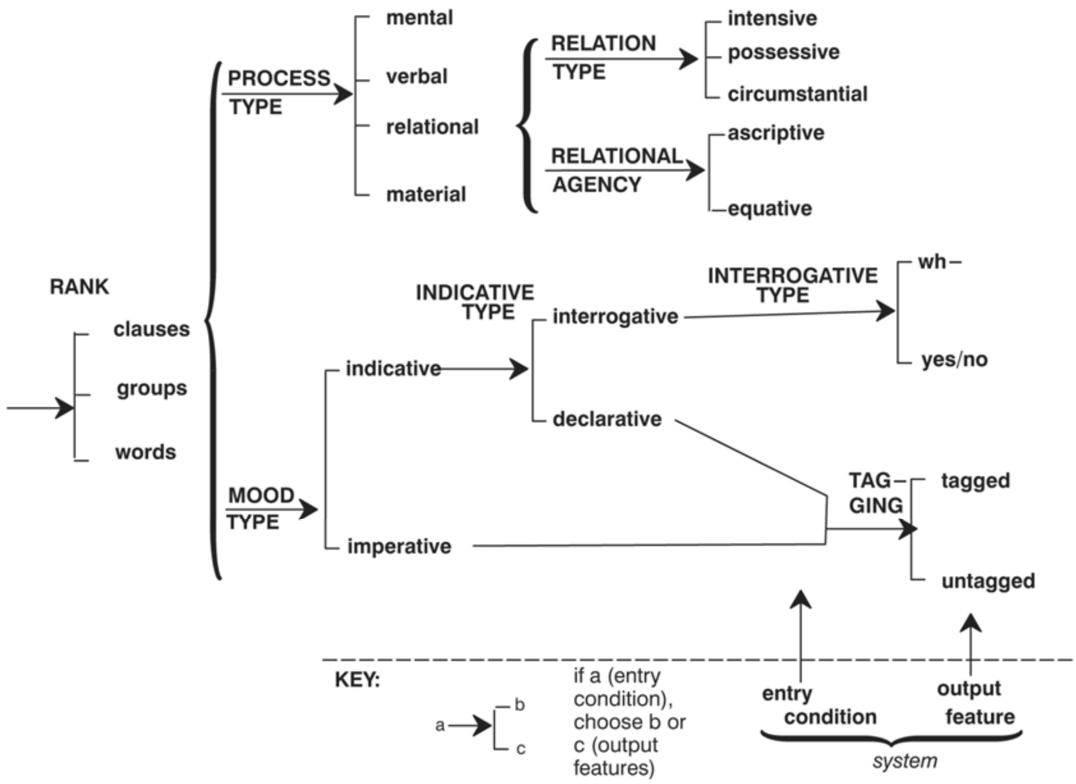


Figure 7: An example system network fragment taken from English grammar

one choice from RELATIONAL AGENCY). Sometimes more than one path may lead to a choice – for example, the TAGGING system can be reached by both of the choices 'declarative' and 'imperative'. This means, therefore, the alternation between 'tagged' and 'untagged' is relevant when *either* 'declarative' *or* 'imperative' are selected, but not, for example, when 'interrogative' has been chosen. The connectivity of the network therefore defines a partial ordering of systems from least *delicate* (most general) to most delicate (most specific).

As illustrative examples, possible English clauses compatible with combinations of these grammatical features covered by the network are shown in examples (1a)–(1d). An ungrammatical example corresponding to an inconsistent combination of features (i.e., a selection of features not possible given the interdependencies of the network) is shown in (1e).

(1)	a.	You are going.	{declarative, untagged}
	b.	You are going, aren't you?	{declarative, tagged}
	c.	Go away!	{imperative, untagged}
	d.	Go away, will you?	{imperative, tagged}
	e.	* Are you going, aren't you?	{interrogative}, {tagged}

Thus, to summarise, a linguistic unit receives, for example, a grammatical description by setting out the 'abstract semiotic choices' that would lead to the construction, or *realisation*, of the linguistic unit in question given the network of possibilities provided by the relevant language. A clause in English may then be seen as requiring an abstract choice between 'indicative', 'interrogative' and so on depending on its intended communicative function.

The individual choices in such a network, moreover, can be characterised equivalently from a static perspective as classificatory *features* or *properties* of some unit of analysis and from a dynamic perspective as *options* or *decisions* for constructing units that are notionally available to some speaker, writer, etc. Below we will make use of both the static and dynamic readings of our network: on the one hand, we are classifying components and subregions of comic book and graphic novel pages and, on the other, we are setting out abstract compositional choices or options that can be taken up during page design. The latter perspective then aligns with, for example, Groensteen's description of comics in terms of *mise-en-page*, i.e., how the space of the page is to be filled (e.g., Groensteen 2007 [1999]: 21), and with Chavanne's (2015) adoption of the designer's perspective for segmenting and fragmenting the page.

Just as is the case of the application of classification networks in linguistic description, however, we must emphasise that this must be seen as an *abstract* characterisation describing what is qualitatively distinct in the medium – there is no assumption of psychological reality whereby the choices set out would correspond to actual design choices that a particular artist weighs during the design process. Any such assumption requires considerable empirical study in its own right and raises quite distinct concerns to those that will occupy us here.

To summarise, therefore, this linguistic approach is relevant for our account here in two respects. First, our classification scheme will be set up in precisely the fashion just sketched for grammatical description: that is, particular units of analysis will be identified and the internal organisation of these units will be captured by relating them to features in a classificatory choice network. Second, we are seeking a classificatory system that can be used directly for larger-scale corpus annotation. The dual nature of our choice network as both a hierarchically organised classificatory scheme, and an unfolding set of options to be selected, allows the network to serve as a backbone for an annotation scheme for page layout. Document annotation then proceeds by examining the options available for classification of some unit and selecting among these according to how well that unit satisfies the constraints defined for the options being considered; we briefly describe this process further in Subsection 5.2.4 below and give an example of some prelminary annotation results in Bateman et al. (2016)

5.2 Units of analysis and the top-level composition distinctions

Our annotation scheme for comics layout is thus modelled as a classification network capturing precisely those abstract 'semiotic decisions' that characterise the options available to the process of *mise-en-page*. In our consideration of a broad range of comics, we have isolated several compositional properties that allow us to begin analysis and classification.

Rather than drawing on the somewhat abstract notion of the 'composition space' envisaged by the creative artist suggested by Chavanne (2015), or the often less precise characterizations in terms of physical pages, 'plates' or spreads otherwise commonly found in the comics literature, we take as the starting point of analysis the notion of the 'virtual canvas' as defined in the GeM framework (Bateman 2008: 16). This is the perceptual unit that a reader is confronted with when engaging with a comic book or graphic novel, and can correspond to a single physical page, a double-page spread, or some smaller segment of a physical page within which a 'comic-like' organisation is sufficiently evident to motivate reading it 'as a comic'. Once a reader (or analyst) has decided that the canvas is to be seen as a comic or (part of a) graphic novel, classification according to our network becomes relevant.

5.2.1 Compositional strategies: tiles and layers

The first stage of analysis/classification is to decide which compositional strategies are being employed. Within any single canvas being read as a comic, it is important to recognise that there is no requirement that one single compositional strategy for page layout must be pursued. The implicit assumption that page compositions must receive individual classification labels is one of the reasons why previous accounts have been forced to offer less precise levels of description, such as 'rhetorical use' or 'rhetorical compositions' (cf. Peeters 2007 [1998]; Chavanne 2015). Walsh's (2012: 53) comments cited above concerning variations in panel size, arrangements and so on are also relevant here therefore.

As a first step towards handling such variations and complexities, we examine whether a canvas can be segmented into areas of local relative homogeneity. This is important for reducing the complexity of the classification task overall and is a significant method for gaining analytic control of the extreme variation regularly employed. We have identified two largely independent dimensions of potential structure for this initial segmentation process, summarised by the following two questions:

- has a homogenous composition strategy been employed for the canvas as a whole or are there *visually distinguishable subareas* on the page where different organizational principles appear to be at work?

- have organizational patterns used on the canvas been arranged to give the visual effect of layering 'in depth', i.e., where some panels or groups of panels appear to be placed further or closer to the reader with respect to others along an axis orthogonal to the page?

An example of a page requiring segmentation according to the first dimension is given in Figure 8(a). The upper part of this page exhibits a simple grid-like organisation, a strip of two panels with vertical and horizontal margins. The lower portion of the page adopts a quite different strategy in that three panels with slanting frames are arranged around a central circular panel.[2] As we will see in more detail below, these

[2]Note here that we do *not* address the placement of the speech bubbles in this decision: speech bubbles, thought balloons, etc. can always be placed across the units identified as panels and groups of

two compositional strategies receive quite different features within our classification scheme and so it is beneficial descriptively to segment this page into two subregions, each of which can then be addressed independently of the other.

We also might expect – although this is at this stage a hypothesis for further empirical investigation – that segmentations of this kind may correlate with narrative concerns. In the present example, this is certainly the case: the upper portion of the page presents an establishing scene or episode, while the lower portion of the page then groups events and parts of events together into a coherent whole centering on Superman catching the person who is falling. It is important, however, that we offer *independent* characterisation of the compositional choices and the narrative organisation to support more effective analysis of the possibilities available. That is, when considering whether separate segments should be distinguished, the focus should be on *visual qualities* of the page or spread as a whole: each segment must be perceived as 'hanging together' visually and so separating itself from its neighbours.

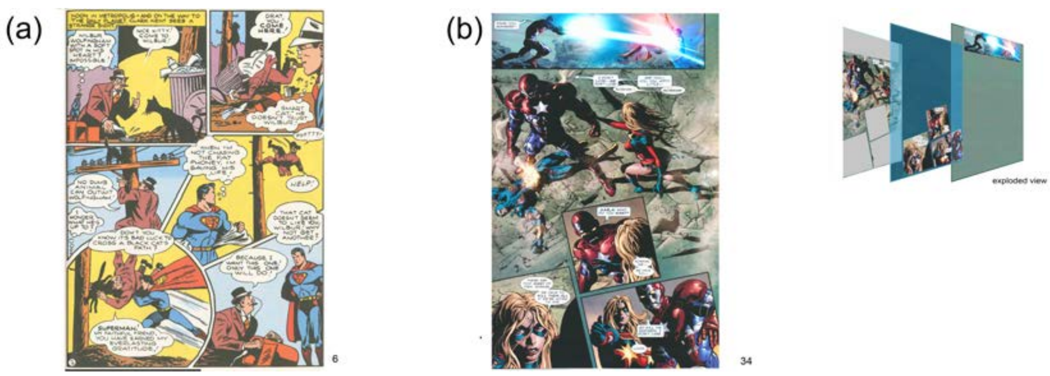


Figure 8: (a) Example of a canvas requiring visual segmentation within a single compositional layer; (b) example of a canvas requiring visual segmentation spread across layers presented 'in depth' as visualised upper right

The second dimension of the initial segmentation involves a very different compositional strategy in that graphical means are used to convey an impression of depth. An example can be seen in Figure 8(b). Such organisations are similar to the previous case in that different composition strategies may be pursued, but is quite different concerning how the distinct strategies are realised simultaneously within the canvas. The previous case places the segments employing distinct strategies either next to one another spatially within a single layer (i.e., in 'parallel') or, as we shall see in a moment, 'embeds' one within another by spatial containment. In contrast, the present case places segments 'on top of' one another within the same visual area.

'In depth' organisations can always be notionally pulled apart into individual layers, as in the 'exploded' view shown here upper right in Figure 8(b). Each layer is then a site for deploying its own compositional strategies. These strategies may differ from

panels. For this reason, it is best to treat them as belonging to a distinct level of annotation. We can then consider further interactions between balloon placement and layout decisions, such as consequences for reading order, without confounding the two variables.

each other, as in Figure 8(b), or they may be the same – for example, we might have a canvas consisting of two grid-like organisations, one placed 'above' the other. As long as sufficient material from the lower layer shows through the gaps in the upper layer, evidence for an organisation in depth remains.

These two basic segmentation distinctions are built into the 'top-most' categories of our classification scheme as follows. We characterize the first distinction as introducing *tilings* of the canvas. Thus the example in Figure 8(a) consists of two tilings, corresponding to the upper and lower portions of the visual unit. The second distinction introduces *layers*. The example in Figure 8(b) consists then of three layers: the panel at the top of the page is clearly 'above' the picture making up most of the page and so is a good candidate for organisation in depth; similarly, the lower portion of the page includes material that is also clearly placed 'above' the central picture – in this case that material is itself a (tilted) collection of further panels, which 'hang together' as a layout unit in their own right. Note that it is also not viable in page (b) to group the two layers made up of the top panel and the bottom panels together into a single layer distinguished internally by tiling because there is no larger unit that would serve to group these prospective tiles together.

This example therefore gives rise to *four distinct units* for classification: the canvas as a whole and three 'sub-units': the top panel by itself, the background image taking up the entire page, and the tilted panel structure at the bottom of the page. Each analysis unit then receives a separate classification of its own in a manner identical to the multiple classifications of hierarchical grammatical units found in grammar.

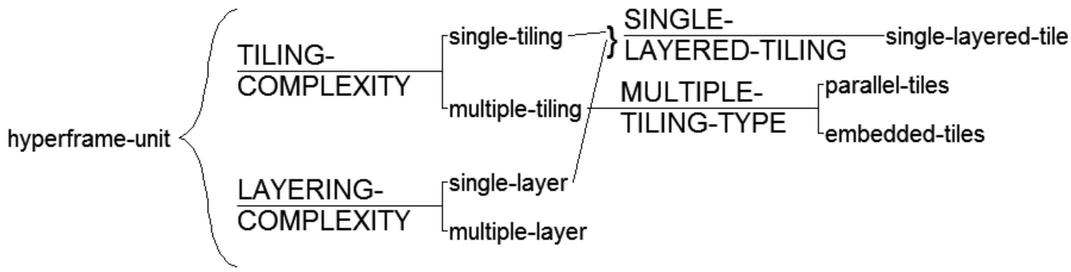


Figure 9: Top-level classification distinctions in the comics layout network

5.2.2 Classifying analytic units

The introduction of tilings and layers into an analysis is regulated by the first portion of our overall layout classification network, which is shown in Figure 9. Here we can see that the grpahic conventions employed are precisely those introduced with th example for grammatical description in FIgure 7 above. To summarise once more, the network shows the abstract choices that are available and the dependencies holding between choices. Choices that depend on previous choices are joined by connecting lines. Rightward facing curly braces indicate that *all* of the more specific choices lying towards the righthand side of the network need to be explored. Conversely, leftward facing braces indicate that all of the less fine-grained connected features lying to the left of the choice point need to have been selected for the designated choice to be possible. Thus, the first distinctions to be drawn here involve 'simultaneous' decisions concerning tiling and layering complexity. Each of these choices consists of two possibilities, either the tiling or the layering involved is 'simple' (i.e., there is only one tiling or layer) or 'complex' (i.e., there are several tilings or layers).

Although we do not show this here in the network explicitly, both tiling and layering distinctions then operate recursively, supporting arbitrarily complex decompositions. Thus, tilings may contain other tilings, and any of these structural segmentations may involve layering. In practice, however, such recursive decompositions appear quite limited. For the multiple levels of Figure 8(b), we have two consecutive selections of 'multiple-layer', calling for further recursive traversal of the network, followed by a 'single-layer' selection terminating the process. Within each of the two recursive traversals, the material on the page leads us to select 'single-layer' again. All three selections of 'single-layer' then introduce a layer for subsequent classification and, for each of these, only the 'single-tiling' choice is motivated. This yields three successive selections of *single-layered-tile*, the rightmost feature shown in this top-level set of classifications, and so traversal is completed.

We proceed to the further, more fine-grained classification of composition choices only after reaching the end of this initial segmentation process – that is, when we have segmented the canvas sufficiently that we are dealing with a collection of single layered tiles. We term these structural units 'layout base units'. Regions of a page or spread that have the internal complexity described by this network correspond closely to the multiple panel units that Groensteen (2007 [1999]: 30) terms *hyperframes*. We therefore adopt this overall name for this portion of our classification network as suggested in the figure.

Figure 10: A further more complex example of multiple tiling

For further illustrative purposes, a more complex example of multiple tiling is shown in Figure 10. Here there is an upper portion of the page, which adopts a composition strategy of a single larger panel in the background and a series of smaller insets running across that panel horizontally, and a lower portion of the page, which is a rather more common grid-like organisation. This means that we again have two tilings, each conforming to one of our major layout classifications that we will see below. The central issue of providing clear criteria for deciding between potentially applicable classifications is already raised by this example, since it would also be possible 'in theory' to propose a characterisation where there is a single grid for the entire page with some insets – this would, however, lose information concerning the locally-restricted position of the top line of insets and so would be dispreferred. We take up this question of ranking possible classifications further below.

5.2.3 Embedded tilings

As shown in the classification network above, there are also two further 'decisions' that need to be made when classifying tilings on a page: tiles may either be placed on a page 'next' to each other in spatial contiguity – as in Figures 8 and 10 – or be *embedded* within other tiles. Embedding within other tiles appears to be relatively rare in comics in general, although frequent in special cases such as McCloud's (1994) description of comics expressed from within the comics form itself. The particular communicative goal that McCloud pursues, which involves considerable discussion of comics strategies, naturally leads to visual 'citations' in which comics layouts appear as examples within the framing comics layout of McCloud's text. Other cases involve 'mental' or 'verbal' citations where, for example, an extended thought or speech balloon may include an entire tiling structure of its own; an example of this kind can be seen in Figure 11.

A superficially similar form of embedding that we filter out of the current annotation layer, however, is when a page or spread exhibits *hierarchies* of framing – that is, there may be larger frames or gutters between groups of panels, which are then themselves separated by smaller frames or gutters. A good example of this can be seen in the original page shown on the left of Figure 12, taken from *War on Crime* from 1937. Here we see that there is a relatively thick gutter running down the centre of the page while the horizontal frame lines are all similar. This gives an overall regular table organisation (see Section 6.1 below). Some of the 'cells' in this table are, however, further divided into panels by much thinner vertical frame lines. This difference in thickness establishes a hierarchy of framing, represented graphically in the centre of the figure. For current purposes, we ignore this further degree of segmentation at the level of tiling and pick it up below as a subclassification of types of grids. Thus, for the purposes of tiling and subsequent grid classification, we treat such cases as if they were structured as shown on the right of the figure – below we will see how the annotation scheme nevertheless preserves the information that more is happening in these layouts than the simple table grid without the recursive organisation.

There are also cases where relatively complex grid patterns can themselves give rise to visual segmentations that may suggest distinct tilings, even though the grids themselves are perfectly regular. Particularly for older comics, it can be unclear whether these visual patterns were deliberately designed or whether they emerge from combinations of independent choices for framing panels. For this reason, we avoid for the present assigning these particular visual segmentations to individual tilings unless

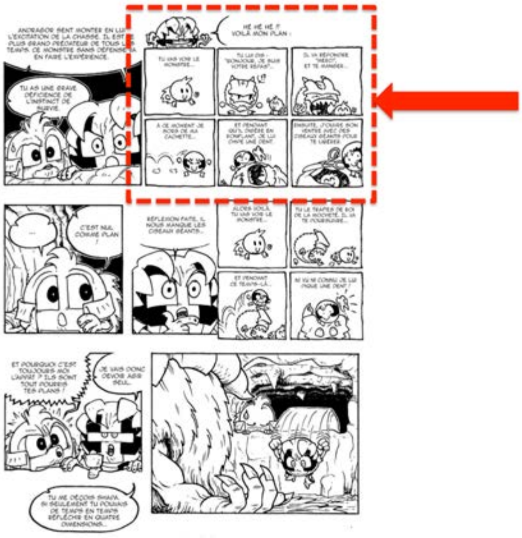


Figure 11: Recursive comics layout example – *Bubble Gôm Gôm*, Cyborg-07, 2014, chapter 2, p. 11

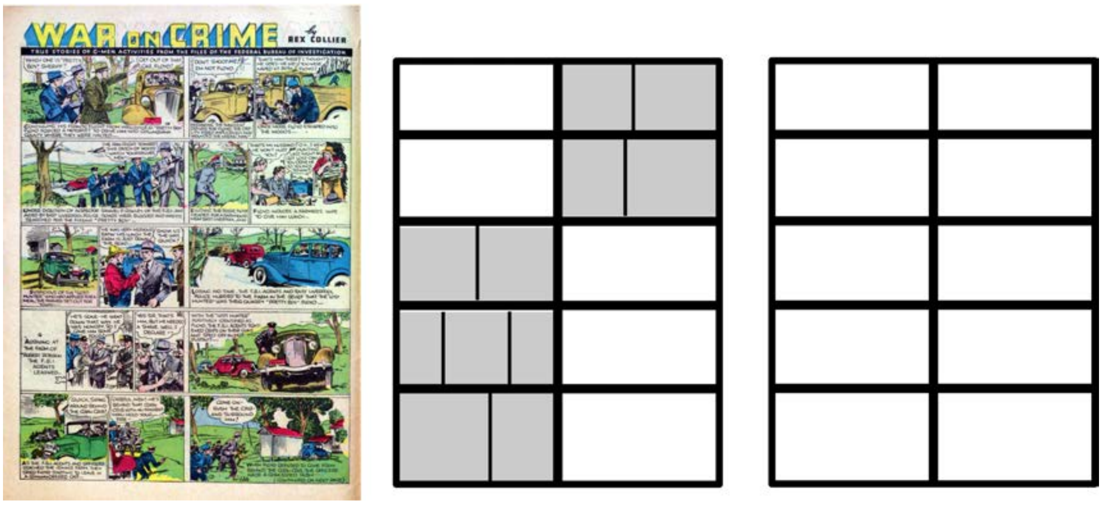


Figure 12: Recursive comics layout example within panels – *War on Crime* (Rex Collier, 1937)

there is clear supporting perceptual evidence and prefer to characterise them in terms of the grid-like distinctions we introduce below.

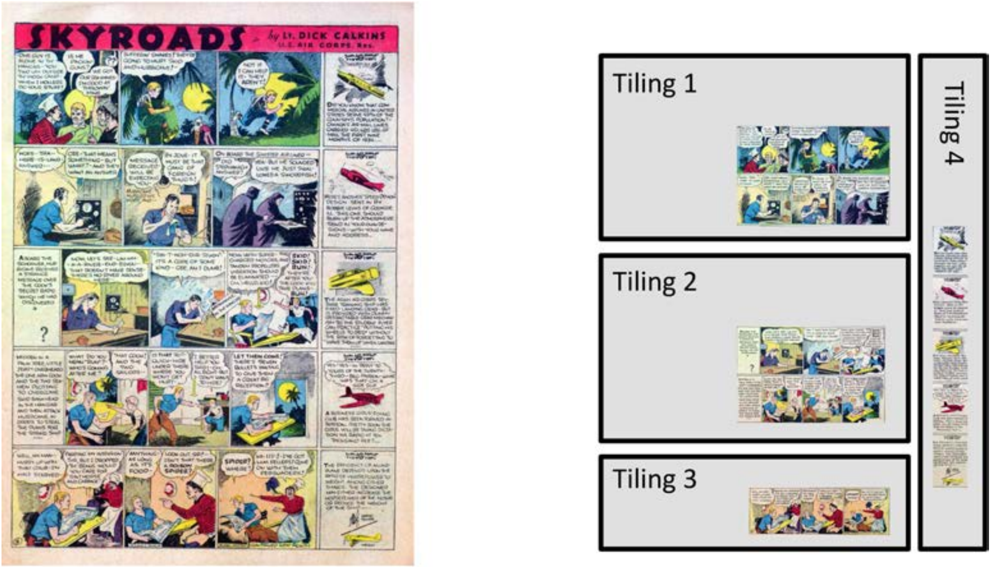


Figure 13: An example of possible multiple tilings created within a regular grid – *Skyroads* (Lt. Dick Calkins, 1937)

When visual segmentations are also supported by strong perceptual features, such as colour or texture, then this may be sufficient to motivate allocating distinguished units to distinct tilings, however. A good example of this is shown in the case illustrated on the left of Figure 13, a page again taken from 1937. Although the main component of this page could be described in terms of a grid-like structure – in fact, what we will term a 'vario-table' below because of a regular table-like organisation with varying numbers of cells per row – the visual similarities and differences among the panels suggest larger scale groupings that would be lost in such a classification. For this reason, the suggested segmentation of a page such as this is as shown on the right of the figure. This respects the fact that the righthand column on the page exhibits very different visual properties compared to the rest of the page and also has a strong inter-panel visual consistency. This is then sufficient to warrant treating it as a tiling in its own right. The fact that the panels within these groups themselves show consistent framing *internally* but not across groups is again considered sufficiently perceptually salient as to warrant tiling. Further empirical experiments involving perception studies may be needed to refine this direction of annotation further.

The two decisions involving tiling and layering are related to one another in interesting ways. On the one hand, when regular patterns or shapes obscure others, then these will generally be perceived as being 'nearer' the viewer than the ones obscured. This involves entire organised areas of the page that are placed in space in front or behind other entire patterns of panels. On the other hand, depth effects may be simultaneously introduced with respect to collections of *differing* tilings — that is, a layer that suggests that it is a unified organisational pattern may stand 'in front of' an arbitrary number of distinct tilings in the 'background'. This is possible when a layer allows sufficient of its background to 'show through' and that background spans more than a single tiling area. In such cases, the tiling decision must be prior. If a layering is created with respect to single tilings, then distinct layers may be created for each such tiling and these layers may well be ordered differently in depth with respect to each other. In cases such as these, in contrast, the layering decision can usefully be made as prior. Otherwise, the layers would not be relative to tilings and would be 'free floating' with respect to the page in a manner in all likelihood not correlating with perception. This shows that there is no logically unique ordering of the decisions with respect to one another.

Finally, it also needs to be emphasised that, as with all of the decisions described below, we attempt as far as possible to *ignore the actual content of the panels*. The annotation level we are concerned with here addresses solely the compositional organisation of the pages analysed, not their content. As Groensteen (2007 [1999]), Cohn (2013b) and others argue, comic page layout does appear to make its own contribution to the reception and understanding of comics, and so it is important to be able to describe this independently in order to pursue empirical research of this contribution. The interaction with content clearly raises a host of further empirical research issues of its own.

5.2.4 Performing annotation

Segmenting pages prior to further annotation is central for subsequent analysis because only then can classification proceed with respect to a homogenous and well motivated segment of the data. If segments were to be mixed, then it would appear as if multiple annotation classifications applied simultaneously and important detail would be lost. For simple pages or spreads, e.g., layouts consisting of single grids of various kinds, the segmentation of the page is simply the whole grid and so the structural units required are trivial. Nevertheless, we still need to take the explicit step of deciding on the segmentation units and identifying these no matter how simple in order that annotation results remain comparable and consistent over an entire data set. In this respect, therefore, there is no difference between simple pages and pages which are complex in terms of tiling and layering.

We show an example of this initial step in Figure 14 with reference to one of the pages in our baseline testset. First we consider the original page as reproduced on the left of the figure and ask which tiling strategy or strategies are being employed. For this decision we do not consider the individual contents of the panels, although broad similarities in framing, in color or composition and so on – i.e., perceptual features that would play a role in grouping regardless of content – may also contribute at this stage. For the current example, we have three panels running horizontally at the top of the page, a central image extending behind those and other panels, and a further panel lower right. This is summarised in schematic form in the middle of the figure. The panel lower right appears to be placed above the image behind and furthermore echoes various visual properties of that image, including colour balance and a depicted face: it is therefore most likely to be considered in a dependent relation with the panel behind

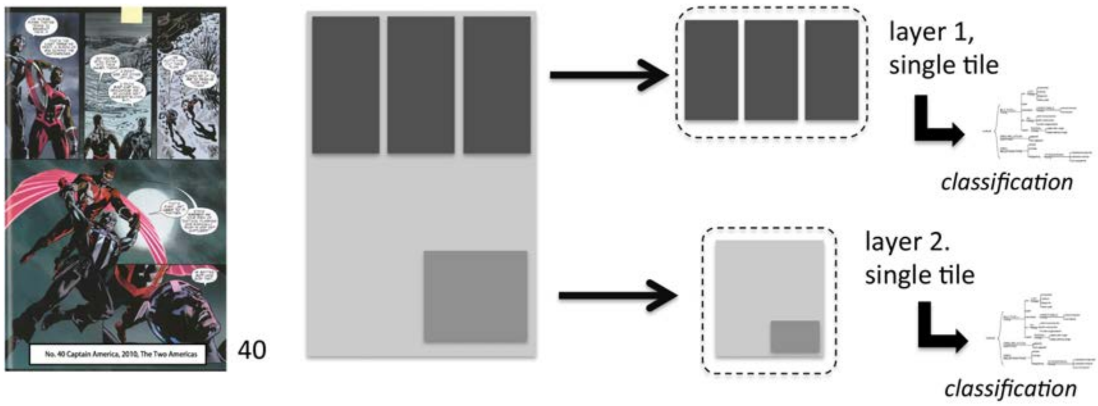


Figure 14: Annotation procedure illustrated with respect to a single example

and is consequently an 'inset' (cf. Section 7.1 below).

The results could also be placed in a table – Excel or similar – which includes separate rows *for each unit analysed*. This means that each tile, each layer, etc. receives its own row. These rows should be named so that they are identifiable. The row then contains a simple list of all the features selected for the unit analysed. Although such an annotation strategy may appear more straightforward, it makes it more difficult to utilise the hierarchical organisation of the classification network and coders will quite likely be required to make many more decisions than actually necessary.

6 The major layout tiling distinctions

Once we have narrowed our analytic focus to a single tiling (perhaps amongst many) within a single layer (perhaps amongst many), then we consider the layout decisions employed solely within that tiling. The next section of our classification addresses qualitative distinctions concerning the use of page space in tiling. Tiling strategies are described from several broad perspectives which together cross-classify the available possibilities. The first large area of decisions we group under COMPOSITION (Subsection 6.1); a second then concerns itself with overall grid properties such as GRID TILTING (Subsection 6.3) and GRID FILLING (Subsection 6.4); a third discusses possibilities created by 'omitting' portions of a grid, i.e., GAPPING (Subsection 6.5); and, finally, the fourth sets out the kinds of FRAMING that a grid may employ (Subsection 6.7).

6.1 COMPOSITION

The first distinctions under COMPOSITION concern whether we are dealing with a single undivided area of the tiling we are addressing or an area that exhibits subdivisions. In order to maintain as much generality as possible, we do not at this point restrict the nature of the elements involved in such subdivisions and use the more neutral term *cell*. Cells are most often panels but may also include blocks of text (e.g., unframed ‘captions’, titles, etc.) or elements which themselves exhibit internal structure, such as embedded tiles. These are then classified according to whether they only contain words and written language (as sometimes occurs in titles), regular panels (including the usual mixture of pictorial, verbal and graphical material), or ‘other’ – what further categories may be useful is not yet known. Whenever we have more than one panel to consider, we need to decide on their composition strategy.

Most of the descriptive challenge for our current purposes then naturally arises from the multiple cell condition where there are layout decisions to be made and a composition strategy must be selected. We distinguish two broad cases.

The first corresponds to a process that Groensteen (2007 [1999]: 41) refers to as *gridding*. This is where some kind of ‘grid’ structure – although we shall see in a moment that this can become quite complex in its own right – is imposed on a tiling. Cells then emerge by virtue of this organizational structure. The second case is where cells, again most typically panels, occur as autonomous units placed within some available canvas space rather than being defined by a grid structure; we term this case *panelling*. Whereas within a ‘gridding’ composition, it is in a sense the grid that creates the cells, in a ‘panelling’ composition it is the spatial distribution of cells that creates any larger patterns discernable. The difference between these is therefore the extent to which an overarching grid is perceptually prominent. When the cells analysed appear to have been placed independently of a grid structure, this is considered in terms of ‘panelling’ rather than ‘gridding’. As we shall see below, these composition strategies are qualitatively very different from one another and so drawing the distinction allows us to bring out both the different options applicable in the two cases as well as some of the commonalities.

Many features describing the spatial distribution of cells are, however, applicable to both gridding and paneling systems. We begin with these shared features and then go on to the separate possibilities offered by gridding and panelling respectively, before briefly setting out some further available options.

6.1.1 MULTICELL TYPE

The classification options under multicell articulate most of the medium-specific variations over patterns of panels on a page that regularly appear within comics. These range in complexity from the simplest ‘table’ or ‘waffle iron’ layouts, with largely uniform panel sizes, to more irregular, but still grid-like, layouts, to free variations motivated by a variety of other factors. Three general types of composition strategy can be distinguished according to whether groups of panels or grids are formed with predominantly straight, vertical and horizontal lines (‘linear’), whether panels are organised in more ‘radial’ patterns, or whether the motivation for cell layout is drawn from other sources. This latter type is only available for panelling options and so we will return to this below. The portion of our classification network covering the multicell configurations possible is shown in Figure 15, together with illustrative layout ‘realisations’ for the main features.

Under the ‘linear’ category, several qualitatively distinct forms of multicell organisations are included that have not, or only partially, been distinguished as such in

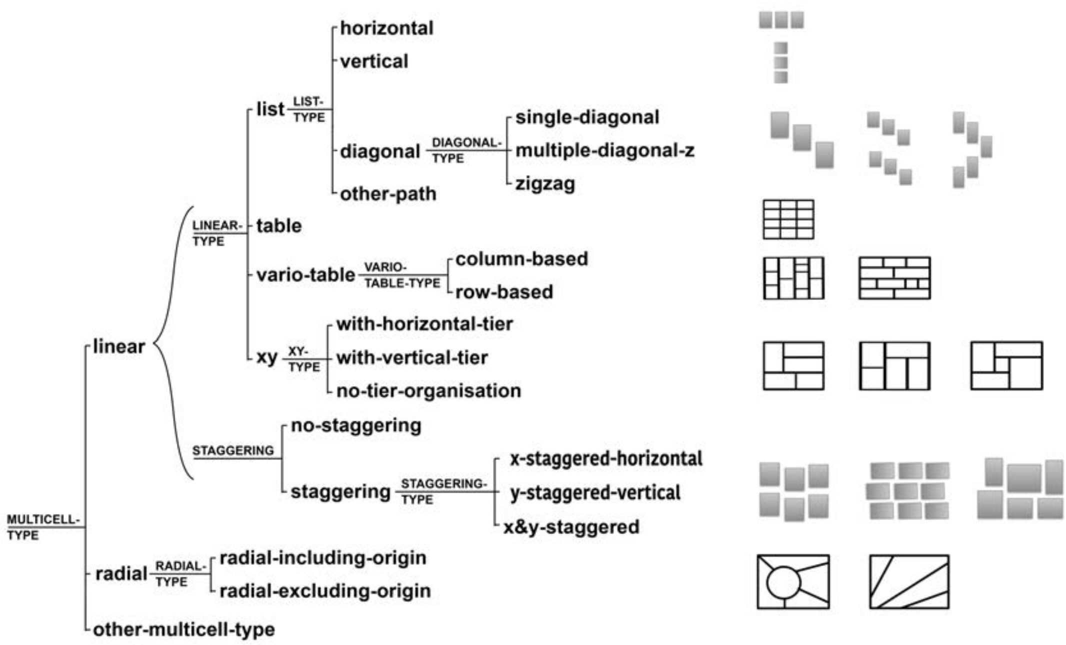


Figure 15: Classification options under 'multicell' configurations with example layout realisations for the main features

previous accounts. Our description treats these in terms of increasing generalisation – that is, each category removes constraints that hold for the category preceding. This also emphasizes a general rule that we apply in the *use* of the classification system: since formally one can always represent an occurrence in terms of a more general category, it is necessary when describing data with this classification system to choose the *most specific* category that applies. This will become clearer as the categories are defined.

The first two categories are straightforward. First, we have the *list*, which indicates that panels are subordinated to a one-dimensional organisation with a clear 'preceding' and 'following' relationship definable between each pair of contiguous panels. This corresponds to the 'strip' emphasised by many comics researchers. At this point, however, we also make explicit the *directionality* of the strip within the broader two-dimensional extent of the spatial region it occupies. Thus, this one-dimensional structure may be spatially oriented vertically, horizontally, diagonally, in a zigzag where the direction changes along the path, or in some other one-dimensional form ('other-path'), such as following a snakelike or other more irregular shape. In terms of an XY structure, lists are compositions corresponding to trees of depth 1.

The second category removes the constraint that we have only one-dimension, thus yielding the *table*. This covers the most restricted form of 'waffle iron' grid common in discussions of comics and so corresponds, for example, to the most restricted form introduced by Chavanne (2015), the 'regular composition'. Tables are completely specified by their fixed numbers and sizes of rows and columns and so exhibit unchanging internal structure along both horizontal and vertical dimensions. Since tables are already two dimensional, the further possibilities available under 'list' of verticality, horizontality, etc. no longer apply. In terms of an XY structure, tables are compositions corresponding to trees of depth 2 in which all subtrees have the same number of child nodes.

The next two linear multicell forms begin to allow more flexibility and are far more common in comics of all types. The third is a further generalisation from tables that relaxes the constraint that subtrees in an XY-structure all have the same number of child nodes. Because of this variability, we label the corresponding category a variable table, or *vario-table*. Thus, whereas in tables it is always possible to identify both rows and columns, in a vario-table it is only possible to identify *one* of these – vario-tables are consequently either row-based or column-based, but cannot be both (since then they would be tables). This form corresponds formally to Chavanne's (2015) 'semi-regular composition'.

The two pages appearing in Figure 5 above, for example, were both row-based vario-tables: the lefthand page having three tiers of 2, 3 and 2 panels respectively, and the righthand page three tiers of 3, 2 and 2 panels respectively. This information can be read off the lowest line of elements in the XY-tree shown in the lower right of the figure. We will occasionally use the shorthand notation for describing vario-tables made up of the number of rows (or columns) per column (or row) depending on whether the table is column-based or row-based respectively. Thus, our examples in the figure are labelled by the number sequence (2,3,2) and (3,2,2) respectively; both are row-based vario-tables.

Vario-tables come about in two ways (which may be combined). In one case, either the number of rows within columns (in the 'column-based' variant) or the number of columns per row (in the 'row-based' variant) differs across rows or columns respectively. In the second case, the cells of the rows (or columns) differ substantially in width (or height). Both conditions give rise to cells within a row or column not being aligned with the cells of a neighbouring row or column. Figure 2 illustrated both these distinguishing features of vario-tables: there is non-alignment caused by the different widths of cells and the last row includes more panels than the other rows. Explicitly labelling the column-based and row-based alternatives allows us to pick out the prominent perceptible organisations also often discussed in the comics literature in terms of 'strips' or 'tiers'. This perceptual prominence correlates with the simple XY-structure exhibited by vario-tables: it will always be possible to perceive a row-like or column-like organisation and this will often (but not necessarily) correspond to reading order (cf. Cohn 2013b; Chavanne 2015). In general, vario-tables should *not* give an impression of substantial alignment across rows (or columns) with minor deviations – any such alignments are to be taken as cues more in favour modifications of tables or of the form to be described next.

The fourth and last form is the most general of the linear layouts and simply removes all constraints from the kinds of XY-trees admitted, thus allowing all the flexibility of XY-trees as discussed above. In this case, there is no requirement of rows or columns holding over the organisation as a whole, although particular sub-areas of the layout may exhibit table or vario-table-like structures. An example of a relatively simple XY layout was given in Figure 6 above; this composition cannot be subsumed to a vario-table because of the non-uniformity of the two subtrees given in the XY-structure (i.e., the subtrees corresponding to the cuts (y_1, x_2) and to y_2 respectively). Whereas this form would also be described by Chavanne (2015) as a 'semi-regular composition' (generally with 'fragmentation'), we consider XY layouts to be qualitatively different to vario-tables. This is supported empirically by Cohn's (2013b) finding that readers select reading paths that diverge from the default Z-path significantly more often in the case of XY layouts as defined here in comparison with layouts that we would describe as vario-tables. Collapsing vario-tables and XY layouts into a single category would loose this empirically-motivated qualitative distinction.

Nevertheless, it may still be the case, as with the example in Figure 6, that the spatial organisation of the grid suggests perceptible horizontal or vertical 'banding' in which a collection of panels can be picked out within the overall XY organisation; this is what gives rise to Chavanne's (2015) focus on 'strips', even within more, in his terms, 'fragmented' compositions. We capture this possibility by additionally allowing 'with horizontal tier' and 'with vertical tier' as disjoint subcategories of XY-structures: the example in Figure 6 exhibited horizontal tiering. Tiering always requires that either horizontal or vertical cuts across the entire encompassing layout unit be possible, thereby introducing rows or columns with complex internal structure. If there is no such perceptual prominence created by the grid, then the feature 'no tier organisation' is selected. There are, however, also 'grid-like' compositions that cannot be described even with the full generality of XY-trees: we return to these in Section 6.1.3.

6.1.2 STAGGERING

All of the types described thus far apply to panels or grids where the edges align linearly at least along one dimension, and often both. It is also possible, however, that the edges do *not* align exactly, but may be 'pushed' out of alignment. Cohn (2013c: 93) characterises such situations in terms of 'staggering' and 'blockage'. For example, it is quite possible that within a table, one or more of the table cells is slightly bigger or smaller than the others so that the edges and corners are dislodged. This tends not to effect reading order and may not even be noticed by readers, constituting only a minor hindrance for, or hesitance in, the reading process; Cohn (2013b) provides a detailed overview of previous studies exploring such reading effects for comics and related artefacts more generally. When it is not possible to find a continuation of a gutter at all because of an intervening panel, Cohn talks of 'blockage'; all less extreme cases he groups under 'staggering' (Cohn 2013b: 3).

Even when quite small variations of this kind occur which are not sufficient to change the overall composition classification, they still need to be recorded as qualitatively differing from the more straightforward cases exhibiting good alignment. The distinctions to be drawn here generally refer to the extent to which a grid-like organisation may deviate from the prototypical rectangular cell patterning where panels are the same size and the corners and edges match up in an orderly fashion. We therefore include this possibility in our classification scheme as follows.

When panel edges are slightly misaligned, we follow Cohn (2013b,c) and characterise this as 'staggering'. When the staggering deforms a sequence of panel edges running horizontally, we classify the layout of the encompassing unit as 'x-staggered-horizontal'; vertical deformation we term 'y-staggered-vertical'. When staggering occurs in both dimensions, the layout receives the classification 'x&y-staggered'. However, and in contrast to Cohn, when the deformation is more substantial we no longer place such cases under 'staggering' and choose instead to characterise the corresponding layouts in terms of vario-tables and XY organisation. We would, consequently, also expect such layouts to be more likely to affect reading path behaviour. The most extreme cases of deformation, which Cohn refers to as exhibiting 'blockage', typically occur whenever a non-maximal 'cut' is required vertically or horizontally to build an XY-tree. Cohn also sees 'blockage' similarly to Chavanne's (2015) notion of 'fragmentation'.

In order to help structure discussions further, we suggest drawing a principled distinction here between descriptions of the *layout* – which is the target of our classification scheme – and descriptions of *what readers may do with that layout*, i.e., potential reading paths. We therefore reserve 'staggering' categories for descriptions of the layout and do not talk of 'blockage' in this regard at all. Differing layout configurations may then give rise to blockage as a reader effect – for example, both vario-tables and XY organisations may be found to have such consequences. Further empirical studies will be required to see whether it is possible to reliably establish tighter boundaries between these categories.

6.1.3 XY-NONCONFORMANT LAYOUTS

Despite the generality of the XY organisation, it is also possible to find layouts whose general patterning may appear to be 'XY-like' in their organisation, but for which it is nevertheless difficult – or even formally impossible – to apply the XY tree organisation and, therefore, any of the formally more specific cases of XY handled here in terms of lists, tables and vario-tables as well. In general these more complex layouts have the property that at some level of segmentation, including the first, it is not possible to provide a 'cut' that divides the space into two subregions. Since XY trees apply a divide and conquer strategy to decomposing the space they are analysing, moving on progressively to ever smaller subareas, the lack of a cut marks a point at which no further XY trees can be constructed. If this occurs at the first segmentation, then no substructure can be captured at all. Here we briefly set out the cases where this can occur and how they are handled within our annotation scheme.

One type of layout preventing clean XY division appears quite common in older comics where grids were drawn by hand and so sometimes exhibit interrelations that work against straightforward columns and rows. Examples of such organisations are shown in Figure 16 along with the classifications that they receive according to the categories that we are introducing in this section. Such layouts have in common that some panel is increased in size in either the X or Y direction (or both) so far that it is no longer perceived as a case of staggering (see above), appearing more like a panel placed 'on top of' (although integrated with) the other panels present or as encroaching

vario-table, overfill, vertical, both | vario-table, overfill, vertical, over | vario-table, overfill, vertical, below

Figure 16: Examples of overfilling (*Airboy*, 6(12): 3, 6, 8)

substantially on the space of those other panels. As a consequence, corners of panels appear in the middle of other panels and correspond to the situations properly described by Cohn (2013b) as ‘blockage’ as discussed above.

In cases where the general linear gridding organisation is not otherwise disturbed, so that it is possible to preserve clear hypotheses concerning reading order analogous to other linear cases, then this circumstance will be treated as grid space *overfilling*. In contrast to what we will describe below as grid underfilling, grid overfilling always has consequences for the grid as a whole since it pushes its containing grid out of shape. For this reason, grid overfilling is considered a further refinement of ‘complex-grid-spacing’ (and, more specifically, ‘grid-space-size-variation’) rather than a further option alongside grid underfilling, which does not have consequences for the overall grid. When the overfilling is both vertical and horizontal, this also gives rise to cases of panel ‘overlap’ (cf. Cohn 2013b), as the corners intrude into the neighbouring panel to obscure those other panels’ own corners. Examples of this can be seen in Figure 17; further discussion is given in Subsection 6.6.

A more complex case is, however, when extra panels are introduced ‘in addition’ to that which an XY-grid would predict. An example, again from the Airboy comics from the 1940s (volume 6, number 12), is shown in Figure 18: here we can clearly see that there is no original panel in an original grid that simply outgrows its allocated space as occurs in overfilling: the panel in the middle on the left of the page is simply *additional* to the remainder of the page, which would have been a simple table had the extra panel not appeared. Because of this ‘additional’ nature, we classify such cases as ‘inset-on-same-layer’ (integrated), even though there is none of the more usually expected ‘dependence’ of inset panels on their ‘dominating’ panels. We return to a brief characterisation of our treatment of inset panels below.

The final type of layout resistant to an XY-tree description is a further extension of the previous type in that panels are found that cannot be related back to an XY (or simpler) linear organisation. However, in contrast to that type, these ‘new’ panels take their own complete space within a grid – they do not encroach on other neighbouring

Figure 17: Examples of overfilling in both dimensions giving rise to spatial overlaps, or ‘intrusions’ (*Airboy*, 6(12): 10, 13)

Figure 18: Example of introduction of a panel additionally to that expected within an XY grid (*Airboy*, 6(12): 5)

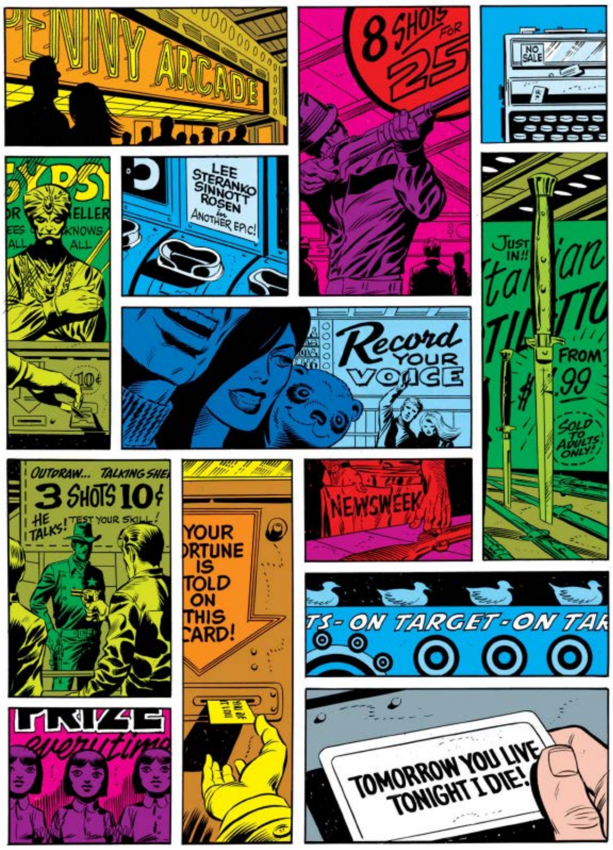


Figure 19: Example of a generally rectangular grid that cannot be given an XY segmentation (page from Jim Steranko and Stan Lee, *Captain America*, #111, Mar. 1969)

panels, leaving corners or other irregularities. This means that the grid as a whole is fully adapted to their appearance leaving a generally rectangular gridding which nevertheless blocks construction of an XY segmentation. Such cases are complex because they may at first glance appear similar to other regular grids; it is only when the search for XY divisions begins that it may be noted that insufficient such divisions can be found. The fully equal status of the panels preventing an XY division also in this case makes any description in terms of deviation or change from a grid inappropriate: no grid space is encroached upon, there is no overlap, and so on. Figure 19 gives an example where this occurs.

In essence, layouts of this kind give rise to 'gaps' in the gridding organisation – they can always be 'repaired' to give an XY-compatible structure by shrinking the gap along either the X or Y dimension (in either direction) and allowing the neighbouring panels to grow accordingly. But this abstract view does not offer an appropriate description of the page for our current purposes because there is no sense in such layouts that the formally present 'gaps' are actually perceived as such – quite the contrary because they appear as full panels just like all the others in the layout. Both Walsh (2012: 53) and Cohn (2013b) discuss the example of this kind shown in Figure 19 and Cohn explicitly noting that it falls outside of his tree description scheme. Cohn's study shows furthermore that such organisations are particularly problematic for the determination of reading paths.

Such XY-blockages can occur at any level in an XY tree, not only at the top-level. This means that at any depth in a layout characterised as an XY tree a leaf node encountered may, in addition to being a single panel, also be a complex XY-resistant region of the kind illustrated. Such regions are, in some respects, then more similar to 'stacked' panels of the kind we introduce below under 'panelling' because a grid is not strictly speaking (i.e., formally) in evidence: however this is often not an intuitive characterisation because the rectangular form of layouts such as Steranko's shown on the right of the figure certainly appear more similar to regular XY grids than panel stacks. And, as Cohn's results show, the areas of indecision concerning reading paths also tend to be isolated to the local areas surrounding the XY 'gaps' – much of the rest of the layout may function more straightforwardly.

For this reason, we maintain a classification for such layouts under the 'linear' portion of our annotation scheme and, within that, under the XY classification. This allows us to make additional classifications according to whether broad horizontal or vertical perceptual organisations are present. To make the distinctions with regular XY segmentations, however, we add the additional parallel classification of either being a 'pure' XY tree or an 'impure' XY. Furthermore, for impure XY layouts, we give an indication of where in an XY segmentation, i.e., how deeply within an XY tree, the XY segmentation must be abandoned. A page such as the Steranko page does not allow any XY cuts to be made at all, and so is impure at level zero; the lefthand page segment in the figure is impure at level one, since it was possible before reaching this segment to apply just one single (horizontal) XY cut. These additions within the XY portion of the network then allow such alternatives to be covered in a manner that permits their encoding and retrieval within the annotation scheme.

What might perhaps be considered a 'transitional' situation is shown in Figure 20, which in certain ways combines features that overlap with those of XY-violations but nevertheless manages to maintain a relatively straightforward grid-organisation. In this case, the 'gap' created by violating an XY organisation is made more visible by virtue of it being filled by an extraposed caption rather than normal panel material. In this case, however, we nevertheless have a composition that is qualitatively far more closely related with regular grid layouts and, in particular, with a straightforward $2x3$ vario-table. This irregularity in the composition comes about by *overfilling* the last panel of the top row (vertically, below) while at the same time *underfilling* the grid space below. This maintains then the shape of the panel lower right without introducing an overlap or *intrusion*. Underfilling also, as usual, opens up space for extraposed captions, which is then the option taken up in the present example. Thus, our analysis of this example would be as a vario-table, with horizontal tiers, plus staggering (see above) and both overfilling (vertical, below) and underfilling, together with extraposed-caption. It is likely that such compositions can usefully receive more specific description, but for present purposes being able to group together layouts where these kinds of decision are made will be sufficient.

As discussed above, smaller variations in gutter alignment that are not sufficient

Figure 20: Example of a combination of underfilling and overfilling (lower portion of page from *Airboy*, 6(12): 7)

to change the overall composition classification are included under the STAGGERING portion of our classification network (see Section 6.1.2). However, in the case of tiering we also need the rows and columns to be sufficiently similar perceptually that they still appear as components of a larger single grid strategy. As with potential vario-tables, if the perceptual distinctions supplied by colour and texture become too great, then we may consider an allocation to distinct tilings as a more appropriate description. These distinctions we draw between tables, vario-tables and XY grids appear to offer useful internal differentiation among cases that in the original GeM-style annotation would all have been simply classified as cases of XY grids without further discrimination.

Figure 21: Examples of radial panel organisation

We then also have a further type within which a combination of geometric organisation and 'list-like' properties combine to form a 'radial' rather than linear organizational principle. These forms are list-like in that they typically impose a clear reading-order (or set of possible reading orders) but differ from the other cases by their inclusion of non-rectangular forms. They are included here by virtue of this strong organisation placed on the reading experience, which is different again from unconstrained geometric forms in general. Layouts of this kind are labelled 'radial' – Figure 21 offers two examples, the lefthand one with an 'origin' visible within the enclosing tile and the righthand one with its origin outside of the enclosing tile and, as a consequence, presumed rather than explicitly visible.

The categories given so far are intended to cover all cases where composition involves the construction of collections of panels that are subordinated to an overarching 'grid'-like organisation. The composition strategy followed involves creating a grid whereby panels follow as necessary consequences. Drawing intersecting horizontal and vertical lines, for example, automatically creates rectangular spaces that may serve as panels. The panels in these 'gridding' compositions are determined by their gridlines: if the lines changes position, then all of the panels subordinate to the grid change. There is also an alternative composition strategy where panels are treated more as autonomous entities in their own right. Such collections of panels may still be perceived as 'unities' or 'wholes' following Gestalt perceptual principles, but are not subordinate to a defining grid structure. We categorise such layouts under a conjunction of the features *other-multicell-type* and *panelling*, as shown in Figure 22.

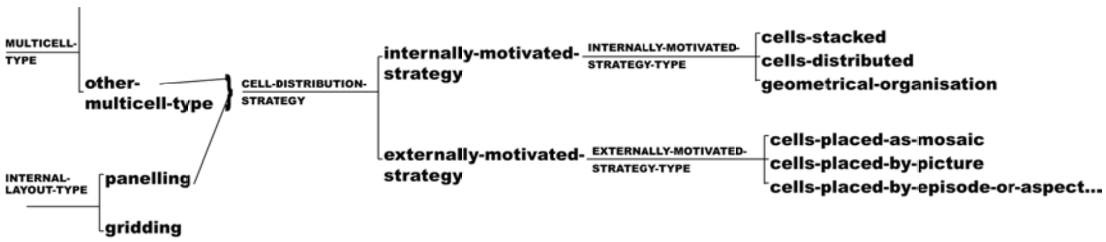


Figure 22: Classification of panelling options for multicell distribution

The possibilities under this category are divided into two groups: 'internally-motivated-strategy' and 'externally-motivated-strategy'. In the 'internal' case, panel placement is still motivated by a logic inherent to the layer in which the panels occur. Three cases are distinguished. *Stacked* cells are when primarily rectangular panels are placed 'as they come', filling a subregion of the tiling – such stacks (or 'heaps' or 'piles') do not need to respect a grid-like organisation and so may be governed by the shape selected for the panels involved. We saw an example of (rotated) cell stacking in the layer at the bottom of Figure 8(b) above – this consisted solely of three panels stacked on top of each other. *Distributed* cells go to the other extreme and include tilings where the panels are distributed over the space as widely as possible. Finally, cells can be placed in order to exhibit a *geometrical-organisation*, where the position of the panels gives an impression of some recognisable geometric form, such as a square, triangle, cross, etc. In all of these cases, creating the composition does not require consideration of

Figure 23: 'Externally' motivated panelling options for multicell configurations – in such cases the motivation for the precise positioning of cells is given by additional constraints *external* to the tile or layer of the cells concerned

information external to the panel layer.

The features under 'externally motivated' operate differently in that creating the composition requires that a further source of motivation for the positions of panels is appealed to from outside of the panel layer itself. The simplest example of such a case is when panels are placed within a spatial region according to some 'underlying' pictorial material quite distinct from the layer in which the panels occur. A straightforward example of this can be seen Figure 23(a): in this case, had the underlying picture of the Joker depicted his arms and legs in slightly different positions, the panels shown would also have had their positions shifted accordingly. This 'external' dependency is then what is picked out by use of the category *by-picture*. Further composition options available here are illustrated in Figure 23(b,c). Panels may be shaped irregularly so that they may be fitted together to tessellate the page space (case (b): *cells-placed-as-mosaic*); or panels may function as elements highlighting 'episodes' depicted in some image 'below' them (case (c): *cells-placed-by-episode*, shown on the lefthand side of the page). Clearly, finer classifications can be pursued here if motivated for some particular data set.

As in many of the cases that we will discuss below, further subclassification here is clearly possible – for example, in terms of the degree and types of regularity of the mosaic pieces themselves – but this has not so far been motivated by the data we have considered and so will not be pursued further here. Also, with respect to our coding or annotation procedures, regular organisations will always be given priority when analysing data. For example, even though it may be possible (and often in fact will be possible) to relate components of a grid or list to 'episodes', we do not pick an episode classification – the episode classification is only employed when this is the *only* way that we can find for relating the panels or areas on the page into a coherent whole.

The primary classifications we have now presented – list, table, vario-table, XY, radial and stacked – characterise in abstract qualitative terms a very high proportion of the layouts we have found in comics so far, although there is, of course, no claim of exhaustivity. Empirical research may lead us to consider other organisations that have not yet been covered adequately. Nevertheless, combined with several remaining further 'modifying' classifications for grid-like layouts in general, we will see that the range of varied layouts covered is already considerable.

When a gridding composition strategy is followed, there are several properties that follow from the fact that a grid as a whole may undergo further deformations that then necessarily have an impact on the panels created within that grid. We describe these specifically grid-based strategies in terms of three cross-classifications that together determine the 'grid layout complexity' of a tiling layout. These three cross-classifications are GRID-SPACING, GRID-TILTING and GRID-FILLING; we describe each in turn in the sections following.

6.2 GRID SPACING COMPLEXITY

First, it is possible for grid-like layouts to show more complex organisations by varying the size and shape of the space they make available for panels in various ways. Any grid thus creates what we will term 'grid space', which is the theoretically available space that they define for filling. Grid space can be made larger or smaller, and can also have varying shapes. Such variation can either be *ad hoc* to meet some requirements of the content to be shown or the narrative pacing, or defined at a more global level by a transformation of the basic grid geometry. The options that the network provides in this area are, first, a choice between 'simple-grid-spacing', where all grid spaces are what would be expected according to the grid regime as a whole, and 'complex-grid-spacing', where there is either some individual variation or systematic deformations of the entire grid. Further choices indicate whether it is size, shape or both size and shape that is varied. The latter case then also admits more extensive geometric transformations of the grid – the most developed area of which involves deformations away from the vertical or horizontal. This is when an entire grid, or individual borders making up the grid, is *tilted*. This can either be independent of size and shape variations, giving simple rotations, or more diverse variations.

More extensive geometric transformations are also possible. In general these differ from the non-grid 'as geometry' classifications shown above in that it is always still possible to see an underlying or related grid organisation – this will mean that by and large the same reading options that were present for a grid will hold. The geometric patterning may, however, introduce further stylistic effects as well as panel groupings and inter-relations of various kinds.

6.3 GRID TILTING

As with the other properties discussed here, the transformations covered involve the overall organisation of the tiling in a grid-like structure and do not refer to the *contents* of panels, which may independently be distorted or depicted at angles for stylistic effect – we consider this the responsibility of distinct annotation layers concerned with

Figure 24: Examples of symmetric tilting and geometric grid transformations

panel and intra-panel characteristics and so are beyond the scope of the present layer in our annotation scheme. Particularly for linear grids, it may be the case that the horizontal lines are tilted, or the vertical lines, or both. Furthermore, such tilting may be uniform for the grid as a whole ('regular' tilting, including rotation) or occur individually ('irregular'). One particular pattern that appears for evident aesthetic effect is when the tilting is performed symmetrically around some axis, so that, for example, verticals are tilted by some angle clockwise one side of the axis and anticlockwise by the same angle on the other side of the axis. In such cases we talk of 'mirrored' tilting and geometric transformation; examples are shown in Figure 24.

Here the leftmost spread shows a 2×4 table with both verticals and horizontals made to follow the lines of the stylised bat shown as a graphic design in the centre of the spread inducing horizontal separation. This is then also an example of complex framing, which we introduce below. The middle example shows a (gapped: see below) 2×6 table where the inner horizontal edges have been swept upwards from the middle of the spread. The last example shows a vario-table where the lower row has both vertical and horizontal transformations, again symmetrically from the centre of the spread. In all of these cases, the grid-like organisation remains, but the shapes of the panels and their framing are systematically altered – either to correspond to more stylised shapes or to fit the content better. Geometric transformations therefore apply when it is still possible to notice that a grid organisation holds but that the shape of the frames employed and how they fit together deviate substantially from the prototypical rectilinear format. The default case remains, however, non-tilted, where the grid frames align with the page orientation.

6.4 GRID FILLING

The presence of a grid also gives rise to further possibilities that can be recognised against the background that the grid provides. For example, since a grid provides 'grid spaces' that are generally occupied by panels, it is possible for the panels provided to fill less than the theoretically available grid space. The limiting case of this is when there is a gap, which we will return to below since it has some special possibilities of its own. 'Grid-underfilling' then allows us to describe the layout situation where certain panels are smaller than the grid makes available. This is distinguished from the case where the grid space itself changes: with grid underfilling the grid remains unchanged, it is just that the space is not fully filled. Such cases of underfilled grid spaces occur quite frequently in combination with captions.

6.5 GAPPING

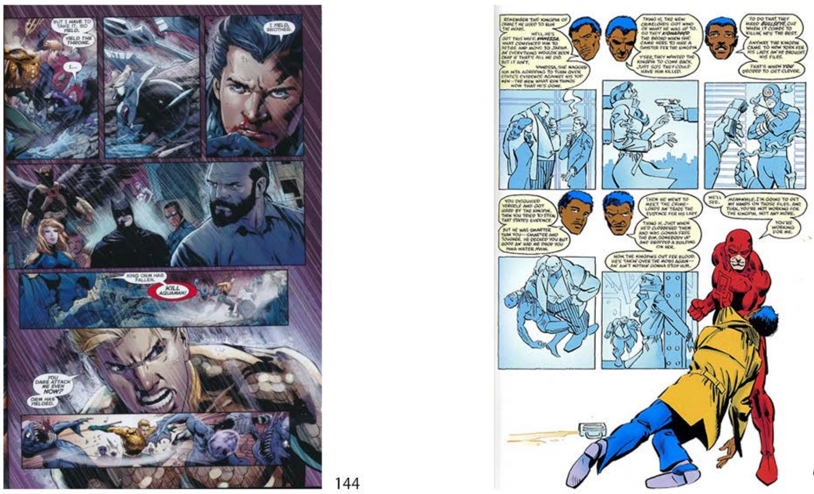
Once an overall grid or panel organisation is apparent it is possible to introduce further variation by *omitting* elements of that structure, which we refer to as 'gapping', or by allowing panel elements within the tiling to *overlap*.

Gapping occurs whenever the layout of the page gives an impression that there is something missing – i.e., there is a 'hole' in the grid organisation. In such cases the feature 'gapped' is selected; otherwise 'non-gapped'. There are then some further options concerned with what happens with the gap – the simplest being that it just remains unfilled, either staying blank or allowing layers lying 'behind' the gapped tiling to show through. This option is then that of 'gap-unfilled'. More interesting alternatives can also use the space opened up by the gap in various ways, however.

One possibility is to provide material that is unframed and which 'bleeds' (see below) beyond what would otherwise be the frame of a panel occupying the gap. This involves the relatively common layout strategy of taking one of the elements of a grid and omitting its frame so as to allow the background of the panel to extend as a background for the grid as a whole. An example of this is shown on the left of Figure 25. In this case, the layout is a vario-table with row organisation (3,1,1,1,1). The element making up the fourth horizontal tier is missing as a panel and instead 'shows through' the gap. Although an alternative might appear to be to see the entire page as being constituted of two layers – a background layer and an unfilled gapped foreground layer – such a representation fails to bring out the fact that the 'background' is not presented independently of the grid structure in the 'foreground'. That 'background' instead provides appropriate content for the vario-table and could easily be replaced by a single framed element. Thus, we classify this example as 'gapped' but with a single layer with a bleeding 'virtual' panel present. The type of gapping involved here is consequently classified as 'filled-same-layer'.

A contrasting example where gapping is also taken to apply but the material filling the gap has a different status is shown on the right of Figure 25. In this case, a regular table organisation (2×3) has a considerable amount missing in its lower righthand corner. The material that is provided there cannot be seen, however, as fitting within a further panel of the form offered by the rest of the grid since it shares almost no visual features; this is taken as a more substantial example of gapping. The material in the corner is then taken as occupying its own layer since it overlaps, and so stands in depth, with respect to the grid. This situation is classified as 'filled-other-layer'. Just how much material may be 'gapped' is not so far specified, although further data analysis may suggest finer classifications – for the present, annotators are recommended to be very conservative in applying the gapping category in order to avoid diverging classifications.

This may lead to a potential indeterminacy, however. For example, if a table were to have sufficient elements missing as to suggest a diagonal or zigzag form, then it



144 66

Figure 25: Contrasting pages with bleeding and 'same layer' filled gapping (left) and 'other layer' filled gapping (right)

is not immediately clear whether a gapped table or an ungapped list is appropriate. In such cases, we take the overall Gestalt impression of the page as determining the classification: only if the grid is sufficiently 'present' perceptually so as to suggest an explicit gap should the gapping option be taken. This then avoids positing tables when there is so little left of the table that it is not perceptually an option. A similar potential indeterminacy can be seen in the example given in Figure 10 below, in which a row of panels is placed over a larger 'background' image. Although in this case it would be possible to see the background image showing through a 'gap' in a table or vario-table, this would require us to imagine a far more complete grid structure than is actually given in the data. For this reason, here too we would not talk of gapping and instead assign different tilings and layers.

6.6 INTRUSION

There are also cases where a panel encroaches on the space of a neighbouring panel; this can occur for a variety of reasons. (Cohn 2013b) refers to this as 'overlap', however we will prefer the term *intrusion* because we often see this configuration in otherwise quite 'flat', or two-dimensional, layouts. It is quite possible for comics that show no features of depth still to employ 'overlapping' panels in a flat, single-layer grid-like tiling (cf. Cohn 2013b). The notion of 'overlap' is not, therefore, so apt as it suggests prescisely this third dimension in which the overlapping can take place.

6.7 FRAMING

Both grid-like and non-grid organisations may make their segmentation visually explicit, for example by leaving white space or background colouring or imagery between elements. If such explicit framing indications are present, we classify the tiling as ‘explicit-framing’. It is also possible, however, to omit such explicit framing and simply to allow visually distinct elements to bump into each other (‘implied-framing’). That is, although framing is always logically present simply by virtue of a grid-like organisation, the question here is whether this is additionally indicated through visual means of its own such as additional separating space or lines between panels or other elements. In the case of explicit framing, then that framing may either be complete, where all the component cells are framed, or partial, where one or more cells do not receive explicit framing.

The case of bleeding introduced above acts together with this latter option. When panels are not explicitly framed, it is possible for their content to spread beyond the theoretically available grid space and to act as a backdrop for neighbouring panels. We distinguish here between two cases: first, where only a single panel ‘bleeds’ into or (most often) behind its neighbours and, second, cases where more than one panel may act in this way. We also classify the layout according to where the panel that is bleeding is placed within, or with respect to, the grid as a whole (e.g., leading, central, trailing or ‘other’), since this is likely also to serve narrative or pacing functions.

Naturally, this can only be applied when the layout itself provides sufficient clues for an ordering relationship. This is generally easier when there is a grid-like structure at work. Thus the examples shown in Figure 25 above, although differing with respect to their gapped status, both exhibit bleeding. The lefthand page is a case of single sourced central bleeding; the righthand page, single sourced trailing bleeding. When deciding on occurrences of bleeding, it is important to consider the lack of presence of the frame – when, for example, the depicted *content* of a panel breaks the frame boundary and impinges on the content of another panel, then this is *not* considered an instance of bleeding. Such cases are picked up at a different layer of annotation to do with panel content if they are minor excursions beyond the frame and within the graphical panel-connecting mechanisms described in Subsection 7.4 below otherwise.

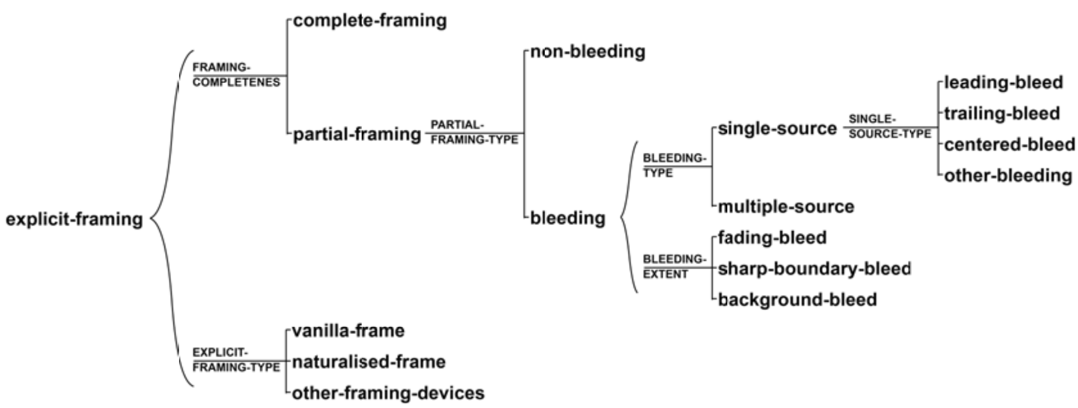


Figure 26: Classification for types of explicit framing

Finally, when present, whether partial or complete, the framings can themselves be more or less complex, ranging from simple lines to quite ornate graphic designs in their own right. The current classification network for this portion of the annotation scheme is shown in Figure 26. Here we can see that it is not necessary for all of the elements constituting the grid to have frames and that the frames may themselves be augmented by graphical resources for stylistic effect, distinguished here as further subclassifications for the type of framing employed. Here we distinguish qualitatively between 'vanilla frames', which are just white space between elements, and framings that use part of the 'diegetic' content of a panel – for example a window frame – as a means of segmenting panels ('naturalised-frame'). It is very likely the case that there are many other distinguishable forms of framing to be found; however, these are all for the present to be classified as 'other-framing-devices'.

Although for the purposes of the present paper we have focused specifically on the macro-organisation of groups of panels as realised through various gridding and similar space allocation strategies, this does not exhaust the range of possibilities for grouping panels in comics and graphic novels. In the section following, therefore, we mention some of these other strategies briefly, showing how they can be added as further layers of annotation building on what we have seen so far.

7 Further visual resources for grouping panels

The kinds of layout organisations that we have considered so far are carried by more or less grid-like or other Gestalt perceptual organisations of entire pages or spreads – i.e., the panels or elements making up those pages or spreads are distributed with a perceptually apparent regularity suggesting that they be grouped in certain configurations rather than others. There are, however, a few other important ways of expressing connections and groupings of panels that are afforded by the spatial distribution of discernable elements on the two-dimensional canvas and which therefore also need to be considered. These strategies all make use of the spatially perceptible features of *connection* and *containment*. These pre-attentive visual features offer resources for grouping and connecting panels of their own. Basic perceptual properties such as these are immediately accessible to readers and function to drive very strong composite configurations of the panels affected. For current purposes, we will describe these possibilities rather briefly in order to characterise their interaction with the options that we have seen so far. There is considerably more here to be researched, preferably within the context of extensive empirical studies.

7.1 Insets

Panel insets operate on the principle of visual containment. Insets are the most common additional panel grouping strategies that we will discuss. The definition and use of inset panels as a further organisational feature of layouts is actually a complex issue in its own right and so we will not discuss it in full here, postponing a more complete account for a paper of its own. However, in brief, we can characterise the inset options as follows, involving both a traditional understanding of inset panels and a somewhat extended usage that appears to warrant further consideration and empirical exploration.

We distinguish first inset panels which appear on the same layer of the described unit and which are integrated into this layer. Following Cohn (2013b: 4), these panels "feature one 'enclosed' panel embedded within another 'dominant' panel." These are then always single panels which are completely surrounded by a further panel, as given, for example, in the lefthand page in Figure 27. Here, the panel in the middle of the page is clearly embedded into and enclosed by the bigger panel surrounding it.

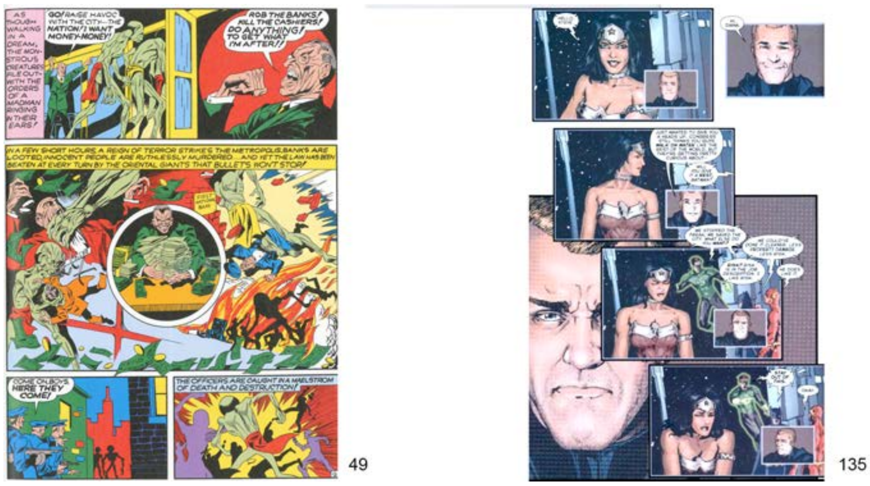


Figure 27: Examples of pages involving Inset Panels

A second type of inset panels is a more specific use of enclosed panels where a repetition or similar reproduction of an inset panel is given as extraposed to the embedding panel, and is therefore situated outside the grid-structure. An example for this is given on the right of Figure 27, where the single panel top right is a repetition of the inset panels given in each of the four bigger panels in the diagonal list. Since it is clearly extraposed to the rest of the panels in this structure and not embedded into a further panel (in the background, for example), we see this as a special case of insetting transitional between a proper embedded inset and an independent panel.

7.2 Subpanel segmentation

Any layout structure used for a tiling may also take the step of including further substructure, or recursive segmentation of its 'panels' into smaller panels. This possibility was introduced and illustrated above in Figure 12. Since rather limited use appears to be made of this possibility, for present purposes we simply mark its presence with a positive feature in the classification system. Further work will be required to know to what extent it will be useful to extend this more finely. In addition, the form of recursive structure may be picked up again in further annotation layers concerned with the 'internal' organisation of panels. Cases such as those shown in Figure 12 suggest that it may be useful to consider 'multiple subpanels' as a particular case of panel organisation; more empirical work is clearly needed. It remains to be seen, for example, just how much of the overall classification network for individual tilings might then apply also to within panel classifications when panels are segmented recursively. Nevertheless, this option is only taken up if there is an explicit visual differentiation between types of frames – substructures created by the use of single frame sizes is covered as usual under vario-tables, XY organisations, etc. as shown above.

7.3 Caption panels

The use of captions in comics and graphic novels is also an area where there is substantially more complexity to tease out, although we will restrict our description here to a basic overview, leaving a detailed account for further papers. Although the basic idea of a caption seems straightforward, the variation that in fact occurs is considerable and has, as with other devices, changed over time. We begin with the basic idea that a caption is more or less readily recognisable as a special form of a framed area on the page that contains text rather than pictorial material and which is not visually/graphically integrated with that pictorial material. We therefore exclude from what we are considering as relevant for the discussion of 'captions' here, for example: speech and thought balloons, onomatopeoic verbal expressions of sounds, as well as 'diegetic' text occurring in the pictorial material.

From this perspective, the simplest kinds of captions are those segmented parts of single panels which contain non-integrated text material and, as essentially *intra-panel* phenomena, these will not be represented within the annotation layer we are discussing in this paper at all – they are to be subsumed within the panels they belong to. The classic form of this was already illustrated in Walsh's CBML example depicted in Figure 3 above, as well as in several of the other pages we have shown so far. All such cases are seen as employing a framed text that is clearly subordinate to its containing panel. Here there are also a range of more complex possibilities: one might consider, for example, whether the panel-external 'talking heads' with speech balloons shown in Figure 25 might be treated as complex captions. In any case, for present purposes their treatment is placed below the level of individual panels and so we will not consider these further here.

Where we do need to say more is when this subordination relation between 'caption' and its panel is no longer so evident and the 'caption' begins to take on a more independent role in the broader layout unit it is contributing to. In this sense, there is a similarity to the increased role of inset panels introduced above – it is not always possible to restrict the account to the simplest cases of containment and subordination. This can have an effect for the grouping of panels within a page layout and so needs to be considered from the perspective of interpanel grouping and compositional organisation as well.

One straightforward way in which this can occur is when the caption is no longer explicitly contained within a panel but instead starts taking its own position within the 'grid space' available. Groensteen refers to some aspects of this phenomenon by pointing that the margin "need not necessarily be virginal" (Groensteen 2007 [1999]: 32). Rather than relying on the notion of 'margin', we consider the 'virtual space' created by particular gridding or tiling decisions. We briefly mentioned cases above where the panels that are presented do not fully fill or occupy this grid space that their surrounding layout 'in principle' makes available; this is one way of using varying frame sizes within a grid-like organisation. In addition, however, there may be cases where the space that is freed by a smaller actual panel is then used for other purposes – for example, for carrying an 'extraposed' caption that is still clearly related to the panel it shares its grid space with but which is visually disconnected from that panel. Such cases are worth distinguishing from the more usual contained panel because it is then relatively straightforward for the extraposed caption to stretch beyond the confines of single subordinated panels so as to include *groups* of panels within its 'scope'. This is one further way of introducing explicit substructure within a page or spread layout. The most appropriate and useful form of annotation for such cases still requires further discussion and empirical analysis, however.

For the present we consequently adopt the simplest position of, first, indicating that the tiling in question has one or more 'extraposed captions' and, second, subtracting the caption from our consideration of other grid options. The panels which then have 'lost space' to the caption are notionally expanded so that a normal grid classification can be made. This temporary solution enables annotators to then return to such cases and consider perhaps a more compelling account on the basis of a broader range of data.

Another interesting direction of expansion that occurs for captions is when they become so independent that they fully occupy an available grid space – that is, they become what visually and graphically can only be seen as full panels in their own right. Panels of this kind can either be made up only of text or may include various additional graphical material. Such cases are often different to normal segmented captions in that the graphic material and the text of the captions are not framed, or segmented, from one another. Here a full continuum of possibilities can be observed, involving increasing independent complexity within the panel functioning primarily as a caption. Several stages on this continuum are illustrated in Figure 28. Of these, only in the last case, where the 'caption' has taken on full panel status of its own, does it appear in the layer of annotation we are concerned with here, since it operates as an ordinary panel with respect to the recognition and classification of an overall grid organisation. Naturally, this also has a structuring function for the narrative or other information that is being given by the panel sequence, but this is not expressed within this level of annotation. In the other cases, we consider these to be further examples of sometimes quite complex *intra-panel* segmentation and so do not provide a separate classification for it here.

7.4 Visually expressed inter-panel dependencies

A further class of groupings of panels is driven by connection: that is, panels or groups of panels distinguished visually may be brought together by explicit 'connectors' that link them on the page. These connectors come in various forms – the simplest being the explicit arrows that were sometimes used in earlier comics when the reading order was unclear. As Cohn (2013b) has demonstrated in recipient studies, typical cases where readers deviate from a simple reading strategy involve occasions where there

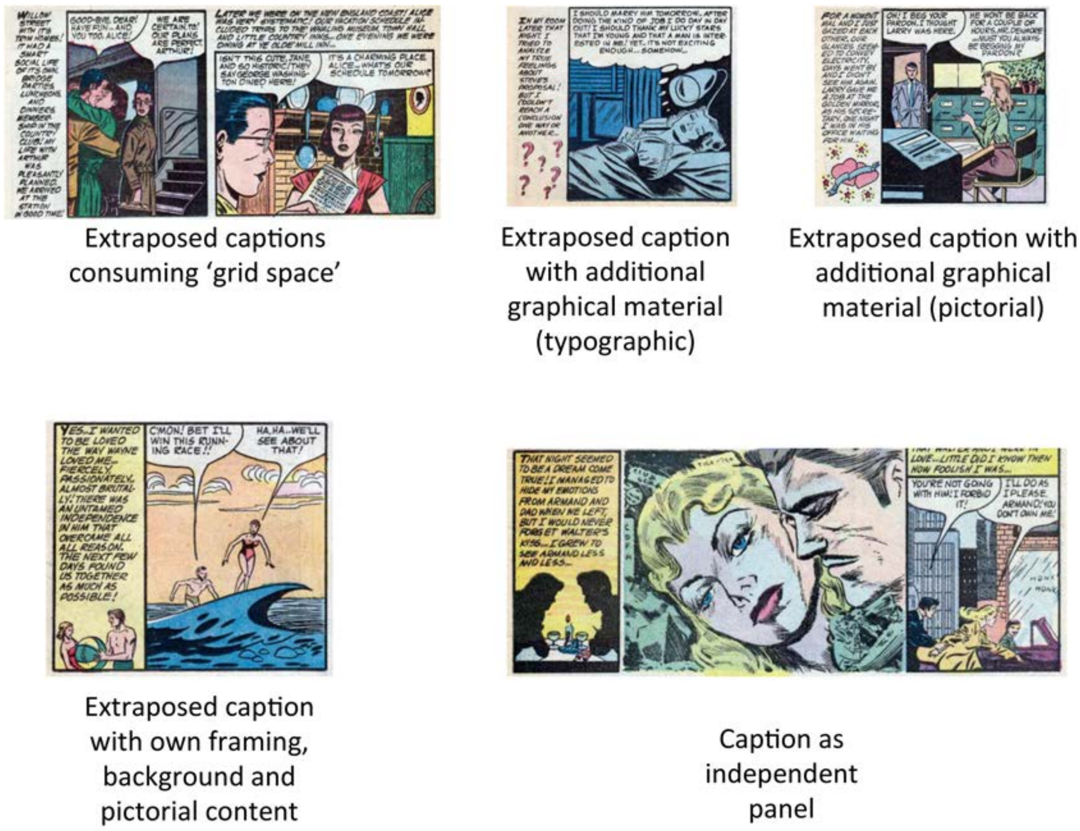


Figure 28: Examples of continuum of panels with extraposed captions – pages taken from *True Romance #12* (1953)

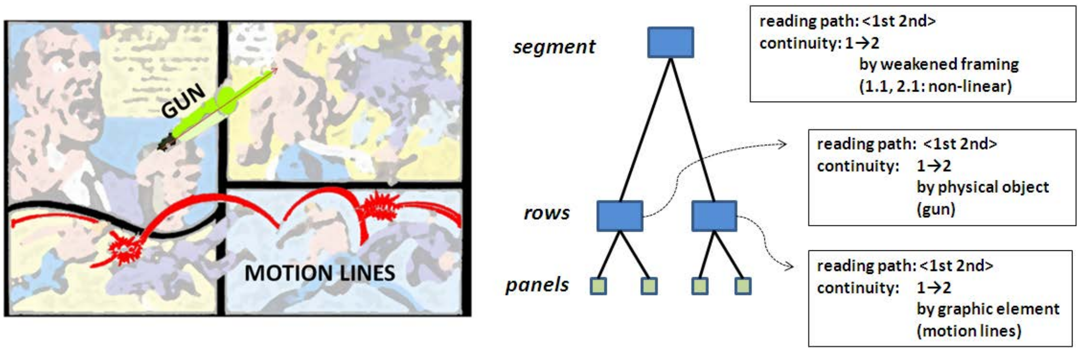


Figure 29: Layout structure annotation with inter-panel effects

is some 'blockage' in the layout where simply following the gutter does not provide sufficient guidance. In our terms above we may find such situations when staggering is sufficiently exaggerated and in all cases of XY-organisations. Early comics designers were evidently aware of the ambiguous reception signals such grid designs introduced and so provided navigation structures via arrows. Nowadays this usage is quite rare as readers are presumed to have the comics literacy required. Moving on from arrows there are, however, a range of further cases where connecting lines of various kinds act as grouping and ordering mechanisms.

Such connections can also be achieved by employing existing graphical technical features in rather different ways. For example, speech balloons may extend beyond the panels of their originating characters and so also link graphically two or more panels. This may also be done using technical features such as movement lines. A relatively sophisticated example of this is shown in Figure 29, which is one of the examples discussed by Walsh. We have modified the page for current purposes to emphasise the graphic connections at issue. These connections are made up of a gun firing across the gutter in the top row and a continuous series of motion lines with sound effects across the bottom row. Walsh remarks concerning the bottom row here that:

> "these graphic moves suggest interesting spatial and temporal juxtapositions and facilitate visual transitions from panel to panel, breaking down the clear separation of narrative moments and investigating a flow approaching (though still very far removed from) the rapid frame-to-frame transitions found in film." (Walsh 2012: 55)

This moves very much in the direction we are following here; there is indeed a connection to be drawn here with film. However, placing this additional information in 'notes' as CBML would suggest is not yet sufficient for supporting empirical research. In order to provide annotations for such connections, therefore, we need to make use of the more fine-grained structure that our annotation scheme supports. Moreover, panel-internal annotation schemes are in any case clearly insufficient here since we are again concerned with groups of panels and larger composition strategies.

Formally this information can be captured as relations expressed at the higher level nodes in layout structures. Here we suggest how this kind of information might be provided more explicitly so that we can search for occurrences and analyse them with respect to other levels of description. In the present case, for example, we can see how the motion lines running across the distinct panels can also be picked up by levels of annotation concerned with ‘actions’ so that both the graphically indicated sounds and the motions can be appropriately captured as single events spanning multiple panels. This is precisely the effect achieved within film by ‘analytical montage’ where single events are similarly decomposed (cf. Bateman and Schmidt 2012: 158–161). Cohn (2013c: 11) makes similar suggestions with respect to comics by separating out the visual composition in terms of panels and a distinct level of ‘events’ with respect to which panels may make varying segmentations. Our annotation here consequently attempts to capture such observations in a way that will support further empirical exploration.

7.5 Intrapanel

Finally, there may also be clear orderings and groupings of panels that are not expressed by explicit gutter framing but *within* larger visuals: such cases should not be distinguished at the panel level at all. They are segmented by the reader/viewer’s interaction with the visual and its content and not by framing cues. This leads to a very different class of possible interpretative strategies, similar to the distinction in film of employing a cut, where events are explicitly broken down, and a so-called *plan sequence*, where several events may be bound together within a single spatio-temporal unity. This overlaps with the previous case and shows once again how several distinct levels of representation are going to be essential for capturing and explaining the phenomena at hand.

8 Transitions and conceptual neighbourhoods

In this section we consider exploratively an important facet of constructing classifications of the kind articulated here for complex aesthetic artefacts; the ideas set out need then to be seen as programmatic and are intended primarily to open up a path of development rather than presenting already established results. The issues raises are, however, sufficiently central for subsequent research that they need already to start receiving more explicit attention. Whereas in the case of linguistic annotation it is common to rely on the ‘digital’ nature of most linguistic distinctions – e.g., a plural form is either a plural form or it is not, a past tense is a past tense and not some category poised between past and present, etc. – the situation with visual media is very different and it is this that needs to be understood when working with annotation procedures and in trying to construct appropriate classification schemes for such artefacts.

The network of options set out in this paper needs can best be seen as imposing a qualitatively motivated structure on what is in fact a theoretically open and continuously variable design space. The features selected are therefore labels for *qualitatively differentiated configurations*. A closer linguistic analogy for the status of the annotation network might then be found not in descriptions of grammatial structures but in descriptions of the vowel space in natural languages. From the perspective of the physical-acoustic properties involved, the vowel space is continuous in two primary dimensions (e.g., openness and front-back). The phonology of any particular language variety will, however, impose further non-continuous structure on that space corresponding to the catalogue of vowel phonemes available in that variety. Distinct (or only partially overlapping) regions within this space may then come to correspond to the 'same' vowel in any particular language varieties: there is no fixed relationship between phoneme and physical-acoustic position. Our annotation network must similarly be seen as imposing a structure on continuous physical dimensions of variation. Lines of demarcation between categories may then depend on a range of features, and this requires empirical investigation.

There is consequently considerable work to be done to investigate possible motivations for our classification network experimentally, particularly with respect to where the boundaries between categories are to be recognized. One set of motivational criteria, for example, may be hypothesised according to perceptual prominence. Another set could be as categories that may be relevant in design or in the conception of the design process: i.e., the features may label design decisions considered by a designer rather than being purely perceptually driven. In this case, the annotation scheme would need to be related to design practice to see if it makes distinctions that correlate with those that designers might draw on. In these respects, therefore, the annotation network as described is a hypothesis concerning the kinds of distinctions that it is beneficial to employ when describing the current state of practice within the medium of comics and graphic novels.

The nature of the relation between continuous material substrates and semiotically imposed structures suggests several ways in which such experimental validation could proceed. We illustrate one of these here that appears particularly promising drawing on an analogy with qualitative spatial representations. Here, although space may be considered to be continuously variable, it has been established that the theoretically available infinite variation can often be usefully abstracted in terms of well behaved qualitative categories. One of the most well known of these is the so-called Region Connection Calculus (RCC, Randell et al. 1992), which characterises the qualitatively different ways in which spatial regions may be related to one another. The simplest Region Connection Calculus is RCC-5, which reduces the theoretically infinite variation to just the five distinguishable cases shown on the lefthand of Figure 30. Thus, regions may either be separate (DR), overlap (PO), contain or be contained in (PP or its inverse, PPi) or equal (EQ). It is possible to reason in terms of these relations in order to calculate 'precise' qualitative relations regardless of their physical manifestation.

Interesting for us here is the notion of *conceptual neighbourhood* (Freksa 1991). If we allow continuous variation in size and distance but at the same time describe the spatial relations obtaining in terms of RCC-5, then the variations can be seen to pass along the fixed network of qualitatively distinguishable situations shown in the right of Figure 30. For example, if we have two spatial regions that are separate (disjoint region: i.e., the leftmost node of the network) and move the regions with respect to each other, then either they move further apart, in which case there is no change in the qualitative RCC-5 situation holding, or they move together so that at some point they

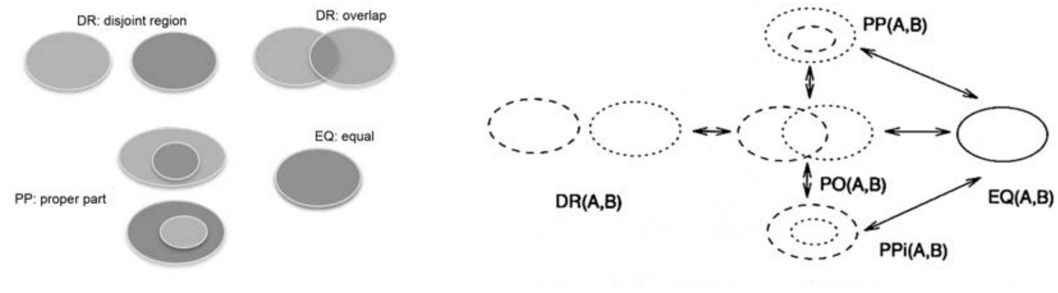


Figure 30: Qualitatively distinguished spatial situations according to RCC-5 and the corresponding conceptual neighbourhood graph

partially overlap. The same transition can occur if we vary the size of one or other of the regions – at some point the situation will change from disjoint region to partial overlap: *no other situation can arise*. This is shown in the conceptual neighbourhood graph by means of the single arrow connecting the disjoint region and partial overlap case. By further variation, three qualitatively distinguishable situations can arise in addition to a return to the starting situation: either the regions become the same size and overlap each completely (equal: rightmost node in the network), or whichever of the regions is smaller may be completely contained within the larger (proper part). The graph in the figure shows all of the transitions that are possible with RCC-5. Thus, given any two situations described in terms of RCC-5, it is possible to determine whether, and which, alternative states needed to be traversed in order to 'move' from one to the other.

Now we consider precisely the same kind of 'transitional' behaviour with respect to our annotation scheme. Given two layouts distinguished qualitatively by being allocated differing features from our network, it should be possible to consider continuous deformations from one layout into the other. Particular boundary conditions can then be isolated that stand as criteria for assigning layouts to one classification rather than another. Let us consider several illustrations of this, since it is quite fundamental to how an annotation scheme for visual media might need to function.

One prominent case is given by the qualitatively distinguished situations of 'insets' and 'multiple layers'. Insets are notionally always at a different layer in the page since they are placed 'in front of' the panel or panels to which they act as insets – however, we restrict the introduction of multiple layers to cases that really demand this complexity. We have suggested above that layering becomes motivated when the material being placed 'in front of' the panel or panels in the background is sufficiently complex in its own right. Since the notion of 'sufficiently complex' is continuous there may be borderline cases that could go in either direction. Thus there may well be transition cases between 'insets' and genuine additional layers of panels.

Consider, for example, the page shown in Figure 31. The classification question here is whether we are dealing with a single panel with several 'inset' panels (cf. Subsection 7.1) or with a background layer with another layer consisting of four panels placed 'on top of' the background. In general, and as we explain further below, insets are seen as necessarily 'dependent' or 'subordinate' to the panels that they occur with

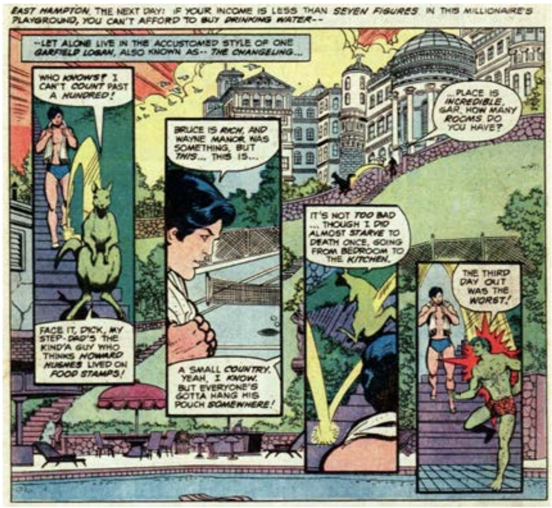


Figure 31: Example comics page with panels placed within a larger panel

– here the four panels running diagonally across the page appear to be far more important in terms of carrying the main development of the narrative at this point. Although it is difficult to assess this purely on the basis of the visual layout without regard to content, the considerable size of the diagonal panels – clearly approximating the norm for the size of full panels elsewhere in this comicbook – suggests that they are more than subordinated additions. For this reason, this case is seen as one involving multiple layers: one background layer made up of a single picture (classified as a cell below) and a layer in front of this containing the four diagonal panels. This analysis is also strengthened by the fact the four panels are readily classifiable in terms of one of our types of grid layout organisations in its own right, independently of the presence or not of the background image. In fact, it may turn out to be most appropriate to consider the presence of single (or double) insets as inherently a transitional case – i.e., a kind of shortcut for introducing full multiple layerings. When the inset is a single panel, it appears unwieldy to consider the full apparatus of multiple layers; however, the more that the inset material takes on a detailed tiling organisation of its own, the more it will be worthwhile to employ the tiling classifications that we introduce below.

Thus, whereas a single panel is a good candidate for an inset, three panels in a grid organisation is a good candidate for layering (rather than three insets). The situation with two panels placed in front of a background panel is then potentially unclear and may go in either direction. For this reason, we need to record the fact that insets and layering are potentially linked via a transition path – here the transition dimension is not simply making regions larger or closer as in the RCC-5 case, but rather that of increasing the complexity and structural integrity of the potentially layered group of panels. Along this dimension we can relate 'insets' and 'multiple layering' as conceptual neighbours. One expected empirical correlate of this connection could be that readers should relate cases of insets and cases of layering as resembling each other more closely than layouts involving completely unrelated features. In this way, empirical support may be successively gathered for the features of the classification network

Figure 32: Example transitional comics page: *Blue Bolt* (1941, vol. 1, #8, p. 5); by Joe Simon and Jack Kirby, published by Novelty Press.

and the relations posited between those features.

Another potential uncertainty arises in the distinction between 'gapped' grids and 'non-gapped' grids. We saw above cases where it may be unclear whether there is a gapped grid or, alternatively, a non-gapped grid with a panel that bleeds beyond its (virtual) panel borders. Here again there needs to be a transition path between, on the one side, non-gapped and bleeding to, on the other side, gapped and layered. Rating page or spread layouts that fall between these two alternative annotations according to resemblance scales may also refine our criteria for judgement as well as making it clearer just which perceptual features readers are attending to in their process of interpretation of the pages encountered.

Finally, consider the page shown in Figure 32. At first glance, this may look like a table, presumably 2×4; then we need to consider the lower portion, where there is what Cohn (2013b) what consider a case of 'blockage' – the gutter running between the lower two panels on the lefthand side does not continue to the right. In our classification, this reader-oriented description is covered by an XY-scheme, so here we would need to assign an XY category with horizontal tiering. However, considered formally, we could also raise the possibility of a vertically organised, i.e., column-based, vario-table. We mentioned above that there needs to be some misalignment of gutters, but we did not specify how much – that is, this is again a place where there may be indeterminacy between category boundaries. We also suggested above that there needs to be sufficient misalignment (what Cohn (2013b) terms 'staggering') to allow a characterisation as a vario-table: this would apply in the present case to disallow the vario-table option. But, again, there may well be cases that we can find that lie between the two categories and where classification may be influenced by further considerations, such as, for example, the design of the panel contents. More study of the interplay of factors at the transition points between categories will be required.

In general, then, we could explore all of the features of our classification network in these terms, gathering data concerning just which layouts are seen as resembling others. This might then be used to feedback into our motivations and criteria for distinguishing features and for arranging them as particular sets of alternatives.

9 Conclusions

In this paper, we have presented an overview of a detailed visually-oriented annotation scheme for comics that is intended to serve as a foundation for empirical multimodal research in the area. We have attempted in our annotation scheme to bring out layout and page organisational patterns which are held to be *distinctively* different from others. We have seen several cases above, where distinctions that may appear similar to others in terms of geometric and other directly observable panel organisation may be considered quite different in intent and effect from the perspective of the comics literate 'reader'. There are also cases where a number of apparently distinct choices are in fact selected as aspects of a single graphical strategy and so need to be captured as such. This has required pursuing a fine line between page organisation seen solely in terms of general organisations of layout of panels and the like and considerations that allow certain aspect of what is being depicted to enter into the discussion. In all cases, however, we have attempted to relate this back to perceptually accessible properties of the material. Thus, although we will avoid consideration of the particular content of panels within page layouts, the fact that, for example, panels may be related visually by virtue of a repetition or variation in a prominent graphical form may very well play a role.

When dealing with creative design, it is clearly not possible, nor desirable, to declare in advance possible or allowable creative choices. It is, however, desirable and beneficial to examine the design options a medium has adopted and to consider both how these change over time and the particular communicative or aesthetic demands and opportunities that have influenced these changes. We suggest that the classification scheme we have presented here constitutes a way of segmenting a potentially continuous design space into qualitatively distinct clusters of properties relevant for discussions of interpretation and effect. Classification according to the features defined can now serve as a tool for pursuing empirically-based corpus studies of the evolving landscape of comics and graphic novels more broadly.

We suggested that the annotation scheme we have presented here is to be seen as a particular structuring of a continuous design space. Within this space, not all options are equal and it is possible to propose qualitatively distinct clusters of features. That is, although it is possible to continuously vary a page, certain perceptual organisations will act as 'attractors' for classification – for example, when possible, a grid-like organisation may be extracted from the page and used by the reader to understand how to interact with the material presented.

This can occur even when the grid-like organisation is considerably ‘deformed’, although at some point the perception of a grid will become so forced that it may no longer be seen. At that point, the page or spread will have crossed over into a different area of the overall classification network. In some cases we have already deliberately drawn distinctions more strongly than may actually turn out to be motivated – the distinction between multiple layers vs. ‘bleeding’ discussed under transitions above being one case in point; we take such cases as hypotheses for direct experimental investigation. There is therefore much still to be done both with respect to exploring borderlines between classification areas and to providing more relevant detail for the classification system itself.

Even with the preliminary stage of classification set out in the present paper, however, we have shown how some straightforward changes across time and design innovations can be tracked. Cross-classifying the annotations across year of production was already seen to reveal significant differences. We can expect many more such differences when turning to cross-classifications involving publishing houses, cultures of origin and, of course, the creative artists involved themselves. One of our own particular concerns is the empirical investigation of the layout forms that are employed for diverse narrative developments and how these have changed over time. Here also a characterisation of the layout options taken up that is largely *independent* of narrative concerns is essential for progress. All such studies have until now been limited by the problem of extracting empirically reliable results from bodies of data. Our comicbooks layout scheme is seen as one contribution towards a new cycle of empirical research for this visually rich medium.

As immediate further developments of this work we need to consider similarly abstract information concerning *intra-panel* technical devices. These have been addressed partly in schemes such as the eBDtheque’s classification of thought and speech balloons (including overall shape and the direction of the ‘tail’ linking balloon with character, etc.), although there is much more that can be usefully captured, such as the shape and form of frames, qualitative spatial descriptions of characters, objects and motions. Such additions again emphasise the importance of both an *open* annotation scheme and a *layered* annotation scheme, so that different kinds of information can be combined from different sources. It would also be interesting to explore the extent to which automated recognition of layout, characters and so on might be brought into the annotation workflow since any reduction in the effort for individual annotation would open up the range of data that can be acquired.

Acknowledgements

We particularly thank the valuable input of Chiaoi Tseng (Bremen) and Carman Ng (Hong Kong) for discussions during development of the annotation scheme as well as to Stephan Packard and group (Freiburg) for detailed feedback and further useful comments, and to Anja Taudte and Christoph Wolf (Bremen) for stylistic corrections and test-driving the framework. Financial support for the work was provided by a funding grant from the German Academic Exchange Service (DAAD, Project 56156404) together with the Hong Kong RGC Joint Research Scheme (Germany/Hong Kong Joint Research Scheme, Project 3-ZG1R), and the Hong Kong RGC General Research Fund Project "Text-image Relations: An Empirical Investigation of the Semiotic Labour of Comicbooks" (B-Q45E). The graphs of classification networks were prepared with Michael O'Donnell's UAM 'Image Annotation Tool' (www.wagsoft.com).

References

Abbott, Lawrence L. 1986. Comic Art: Characteristics and Potentialities of a Narrative Medium. *Journal of Popular Culture* 19(4). 155–176.

Bánski, Piotr. 2010. Why TEI stand-off annotation doesn't quite work: and why you might want to use it nevertheless. In *Proceedings of Balisage: The Markup Conference* (Balisage series on markup technologies 5), doi:10.4242/BalisageVol5.Banski01.

Bateman, John A. 2008. *Multimodality and Genre: a foundation for the systematic analysis of multimodal documents*. Basingstoke: Palgrave Macmillan.

Bateman, John A. 2013. Multimodal Corpus-Based Approaches. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*, Blackwell Publishing Ltd. doi:10.1002/9781405198431.wbeal0812.

Bateman, John A. 2014. Using Multimodal Corpora for Empirical Research. In Carey Jewitt (ed.), *The Routledge Handbook of multimodal analysis*, 238–252. London: Routledge 2nd edn.

Bateman, John A., Judy L. Delin and Renate Henschel. 2004. Multimodality and empiricism: preparing for a corpus-based approach to the study of multimodal meaning-making. In Eija Ventola, Cassily Charles and Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 65–87. Amsterdam: John Benjamins.

Bateman, John A. and Karl-Heinrich Schmidt. 2012. *Multimodal Film Analysis: How Films Mean* Routledge Studies in Multimodality. London: Routledge.

Bateman, John A., Francisco O.D. Veloso, Janina Wildfeuer, Felix HiuLaam Cheung and Nancy Songdan Guo. 2016. An open multilevel classification scheme for the visual layout of comics and graphic novels: motivation and design. *Journal of Digital Scholarship in the Humanities* .

Bateman, John A. and Janina Wildfeuer. 2014. A multimodal discourse theory of visual narrative. *Journal of Pragmatics* 74. 180–218. `http://dx.doi.org/10.1016/j.pragma.2014.10.001`.

Caldwell, Joshua. 2012. Comic panel layout: A Peircean analysis. *Studies in Comics* 2(2). 317–338.

Cesarini, Francesca, Marco Lastri, Simone Marinai and Giovanni Soda. 2001. Encoding of Modified X-Y Trees for Document Classification. *Proceedings of the Sixth International Conference on Document Analysis and Recognition (ICDAR'01)* 1131–1136. `http://doi.ieeecomputersociety.org/10.1109/ICDAR.2001.953962`.

Cesarini, Francesca, Simone Marinai, Giovanni Soda and M. Gori. 1999. Structured Document Segmentation and Representation by the Modified X-Y tree. *Proceedings of the Fifth International Conference on Document Analysis and Recognition (ICDAR'99)* 563–566. `http://doi.ieeecomputersociety.org/10.1109/ICDAR.1999.791850`.

Chavanne, Renaud. 2015. The composition of comics. *European Comic Art* 8(1). 111–144. Translated by Anne Miller.

Cohn, Neil. 2010. The limits of time and transitions: challenges to theories of sequential image comprehension. *Studies in Comics* 1(1). 127–147.

Cohn, Neil. 2013a. Beyond speech balloons and thought bubbles: The integration of text and image. *Semiotica* 197. 35–63.

Cohn, Neil. 2013b. Navigating comics: an empirical and theoretical approach to strategies of reading comic page layouts. *Frontiers in Psychology* 4(186). 1–15. doi:10.3389/fpsyg.2013.00186.

Cohn, Neil. 2013c. *The Visual Language of Comics: Introduction to the structure and cognition of sequential images*. London: Bloomsbury.

Cohn, Neil. 2013d. Visual Narrative Structure. *Cognitive Science* 37(3). 413–452.

DeRose, S.J., David G. Durand, Elli Mylonas and Allen H. Renear. 1990. What is text, really? *Journal of Computing in Higher Education* 1(2). 3–26.

Dipper, Stefanie. 2005. Stand-off representation and exploitation of multi-level linguistic annotation. In *Proceedings of Berliner XML Tage 2005 (BXML 2005)*, 39–50. Humboldt University, Berlin.

Eisner, Will. 1992. *Comics and sequential art*. Princeton, WI: Kitchen Sink Press Inc.

Freksa, C. 1991. Conceptual neighborhood and its role in temporal and spatial reasoning. In M. Singh and L. Travé-Massuyès (eds.), *Decision Support Systems and Qualitative Reasoning*, 181–187. Amsterdam: North-Holland.

Groensteen, Thierry. 2007 [1999]. *The system of comics*. Jackson, Miss.: University Press of Mississippi. Translated by Bart Beaty and Nick Nguyen, from the original French *Système de la bande desinée* (1999).

Guérin, Clément, Christophe Rigaud, Farid Ammar-Boudjelal Antoine Mercier, Karell Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Jean-Marc Ogier and Arnaud Revel. 2013. eBDtheque: a representative database of comics. In *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR)*, .

Halliday, Michael A. K. and Christian M. I. M. Matthiessen. 2004. *An Introduction to Functional Grammar*. London: Edward Arnold 3rd edn.

Hiippala, Tuomo. 2012. The localisation of advertising print media as a multimodal process. In Wendy L. Bowcher (ed.), *Multimodal texts from around the world : Linguistic and cultural insights*, 97–122. Basingstoke: Palgrave Macmillan.

Horstkotte, Silke. 2013. Zooming in and out: panels, frames, sequences, and the building of graphic storyworlds. In Daniel Stein and Jan-Noël Thon (eds.), *From comic strips to graphic novels. Contributions to the theory and history of graphic narrative* (Narratologia / Contributions to Narrative Theory 37), 27–48. Berlin and New York: de Gruyter.

Kannenberg, Jr., Gene. 1993. The comics of Chris Ware: Text, image and visual narrative strategies. In Robin Varnum and Christina T. Gibbons (eds.), *The language of comics. words and images*, 174–197. Jackson: University Press of Mississippi.

Kise, Koichi, Akinori Sato and Motoi Iwata. 1998. Segmentation of Page Images Using the Area Voronoi Diagram. *Computer Vision and Image Understanding* 70(3). 370–382. doi:http://dx.doi.org/10.1006/cviu.1998.0684. `http://www.sciencedirect.com/science/article/pii/S1077314298906841`.

Kress, Gunther and Theo van Leeuwen. 2001. *Multimodal discourse: the modes and media of contemporary communication*. London: Arnold.

Ludwig, Logan. 2015. *Moving panels: translating comics to film* Edwardsville, Illinois. Sequart Organization.

Maderlechner, Gerd, Angela Schreyer and Peter Suda. 2000. Extraction of Relevant Information from Document Images Using Measures of Visual Attention. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, vol. 4, 4385–3489. Washington, DC, USA: IEEE Computer Society. `http://dl.acm.org/citation.cfm?id=876868.877815`.

Magnussen, Anne. 2000. The semiotics of C.S. Peirce as a theoretical framework for the understanding of comics. In Anne Magnussen and Hans-Christian Christiansen (eds.), *Comics & culture: Analytical and theoretical approaches to comics*, 193–207. Copenhagen, Denmark: Museum Tusculanum Press University of Copenhagen.

McCloud, Scott. 1994. *Understanding comics: the invisible art*. New York: HarperPerennial.

Nagy, G. and S. Seth. 1984. Hierarchical representation of optically scanned documents. In *International Conference on Pattern Recognition (ICPR)*, 347–349. IEEE Computer Society.

Namboodiri, Anoop M. and Anil K. Jain. 2007. Document Structure and Layout Analysis. In Bidyut B. Chaudhuri (ed.), *Digital document processing* Advances in Pattern Recognition, 29–48. Berlin, London: Springer. doi:10.1007/978-1-84628-726-8_2.

Ochs, Elinor. 1979. Transcription as Theory. In Elinor Ochs and Bambi B. Schieffelin (eds.), *Developmental Pragmatics*, 43–72. New York: Academic Press.

Peeters, Benoît. 2007 [1998]. Four conceptions of the page. *ImageTexT: Interdisciplinary Comics Studies* 3(3). `http://www.english.ufl.edu/imagetext/archives/v3_3/peeters/`. Translated from: *Case, planche, récit: lire la bande dessinée* (Paris: Casterman, 1998), pp. 41–60. Translated by: Jesse Cohn.

Pose, Javier, Patrice Lopez and Laurent Romary. 2014. A Generic Formalism for Encoding Stand-off annotations in TEI. Tech. Rep. hal-01061548 HAL open archives. `https://hal.archives-ouvertes.fr/`.

Randell, D.A., Z. Cui and A.G. Cohn. 1992. A spatial logic based on regions and connection. In Bernhard Nebel, Charles Rich and William R. Swartout (eds.), *Proceedings of the 3rd. International Conference on Knowledge Representation and Reasoning*, 165–176. San Mateo: Morgan Kaufmann.

Rigaud, Christophe, Clément Guérin, Dimosthenis Karatzas, Jean-Christophe Burie and Jean-Marc Ogier. 2015. Knowledge-driven understanding of images in comic books. *International Journal on Document Analysis and Recognition (IJDAR)* 18(3). 199–221. doi:10.1007/s10032-015-0243-1. `http://dx.doi.org/10.1007/s10032-015-0243-1`.

Rust, Martha. 2008. ”It's a magical world”: the page in comics and medieval manuscripts. *English Language Notes* 46(2). 23–38.

Sumathi, C.P. and N. Priya. 2013. A combined edge-based text region extraction from document images. *International Journal of Advanced Research in Computer Science and Software Engineering* 3(8). 827–835.

Sutheebanjard, Phaisarn and Wichian Premchaiswadi. 2010. A modified recursive x-y cut algorithm for solving block ordering problems. In *2nd International Conference on Computer Engineering and Technology (ICCET)*, vol. 3, 307–311. IEEE. doi:10.1109/ICCET.2010.5485882.

Thomas, Martin. 2009. Developing multimodal texture. In Eija Ventola and Arsenio Jesús Moya Guijarro (eds.), *The world told and the world shown: multisemiotic issues*, Basingstoke: Palgrave Macmillan.

Vanhoutte, Edward. 2004. An Introduction to the TEI and the TEI Consortium. *Literary & Linguistic Computing* 19(1). 9–16. http://dx.doi.org/10.1093/llc/19.1.9.

Walsh, John A. 2012. Comic Book Markup Language: An Introduction and Rationale. *Digital Humanities Quarterly* 6(1). http://digitalhumanities.org/dhq/vol/6/1/000117/000117.html.

Xiao, Yi and Hong Yan. 2003. Text region extraction in a document image based on the Delaunay tessellation. *Pattern Recognition* 36(3). 799–809. doi:http://dx.doi.org/10.1016/S0031-3203(02)00082-1. http://www.sciencedirect.com/science/article/pii/S0031320302000821.

Appendix: Full Layout Classification Network

In this appendix we set out the full layout classification scheme (V1.2) and list changes made with respect to previous versions in order to allow updates in existing annotations more easily. Loading previous annotations in the UAM corpus tool, for example, and making the changes listed here should update the annotations accordingly.

Change list

The following are changes in feature names made since V1.1 (Bateman et al. 2016).

Date	System	Change
2016-06-09	STAGGERING-TYPE	all features made uniformly to refer to *staggering*: 'x-blocked-horizontal' → 'x-staggered-horizontal', 'y-blocked-vertical' → 'y-staggered-vertical'


layout-base-unit
COMPOSITION
multicell
MULTICELL-TYPE
linear
LINEAR-TYPE
list
LIST-TYPE
horizontal
vertical
diagonal
DIAGONAL-TYPE
single-diagonal
multiple-diagonal-z
zigzag
other-path
table
vario-table
VARIO-TABLE-TYPE
column-based
row-based
xy
XY-TYPE
with-horizontal-tier
with-vertical-tier
no-tier-organisation
STAGGERING
no-staggering
staggering
STAGGERING-TYPE
x-staggered-horizontal
y-staggered-vertical
x&y-staggered
radial
RADIAL-TYPE
radial-including-origin
radial-excluding-origin
other-multicell-type
CELL-DISTRIBUTION-STRATEGY
internally-motivated-strategy
INTERNALLY-MOTIVATED-STRATEGY-TYPE
cells-stacked
cells-distributed
geometrical-organisation
externally-motivated-strategy
EXTERNALLY-MOTIVATED-STRATEGY-TYPE
cells-placed-as-mosaic
cells-placed-by-picture
cells-placed-by-episode-or-aspect
INTERNAL-LAYOUT-TYPE
panelling
gridding
GRIDDING-TYPE
simple-grid-spacing
complex-grid-spacing
COMPLEX-GRID-SPACING-TYPE
grid-space-size-variation
grid-space-shape-variation
grid-space-size-shape-variation
GRID-FILLING
grid-filling
grid-underfilling
GRID-TILTING
tilted
VERTICAL-TILTING
no-vertical-tilting
vertical-tilting
VERTICAL-TILTING-TYPE
regular-vertical-tilting
irregular-vertical-tilting
mirrored-vertical-tilting
HORIZONTAL-TILTING
no-horizontal-tilting
horizontal-tiliting
HORIZONTAL-TILTING-TYPE
regular-horizontal-tilting
irregular-horizontal-tilting
mirrored-horizontal-tilting
untilted
INTRUSION
non-intruding
intruding
GAPPING
gapped
GAPPED-TYPE
gap-unfilled
filled-same-layer
filled-other-layer
non-gapped
cell
FRAMING
implied-framing
explicit-framing
FRAMING-COMPLETENES
complete-framing
partial-framing
PARTIAL-FRAMING-TYPE
non-bleeding
bleeding
BLEEDING-TYPE
single-source
SINGLE-SOURCE-TYPE
leading-bleed
trailing-bleed
centered-bleed
other-bleeding
multiple-source
BLEEDING-EXTENT
fading-bleed
sharp-boundary-bleed
background-bleed
EXPLICIT-FRAMING-TYPE
vanilla-frame
naturalised-frame
other-framing-devices
LAYOUT-ROTATION
no-rotation
rotation
ROTATION-TYPE
entire-layout-rotation
individual-element-rotations
INDIVIDUAL-ELEMENT-ROTATIONS-TYPE
regular-elements-rotations
irregular-element-rotations
EXISTENCE-OF-SUBPANEL-STRUCTURE
no-subpanel-segmentation
subpanel-segmentation
EXISTENCE-OF-INSET-PANELS
inset-panel
INSET-PANEL-TYPE
inset-on-same-layer-(integrated)
inset-extraposed
no-inset-panel
EXISTENCE-OF-EXTRAPOSED-CAPTIONS
no-extraposed-captions
extraposed-captions