



The Visual Language Research Corpus (VLRC): an annotated corpus of comics from Asia, Europe, and the United States

Neil Cohn¹ · Bruno Cardoso¹ · Bien Klomberg¹ · Irmak Hacimusaoğlu¹

Accepted: 26 May 2023 / Published online: 14 July 2023
© The Author(s) 2023

Abstract

The Visual Language Research Corpus (VLRC) is a dataset of annotations of 376 stories from comics from the United States, northwestern Europe, and East Asia, along with analysis of the complete 10 year run of the *Calvin and Hobbes* comic strip. Comics were annotated using 254 constructs from Visual Language Theory, a framework focusing on the linguistic and cognitive structures involved in visual and multimodal information. These annotations include analysis of panel's attentional framing structure and filmic shot scale, the situational changes across panels, page layouts, multimodality, visual morphology, and path structure.

Keywords Visual language · Comics · Corpus analysis · Cross-cultural comparison

1 Introduction

As research on comics has grown substantially over the past decades, increasing focus has been given to empirical and computational analyses of comics and visual narratives. Early annotation efforts remained fairly small in scope. For example, in a dissertation, Neff (1977) compares comics from several genres in terms of their panel shapes, angle of viewpoint (lateral, high, low), and filmic shot scale (close, wide). In addition, McCloud (1993) reports on his quantitative, yet less formal, analysis of the changes in meaning across panels in comics from the United States, Japan, and Europe. More recently, larger corpora have been built to analyze comics (for review, see Laubrock & Dunst, 2020), using both researcher annotations and computational methods (Dunst et al., 2017; Fujimoto et al., 2016; Guérin et al., 2013), with fairly

✉ Neil Cohn
neilcohn@visuallanguagelab.com

¹ Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition, Tilburg University, 5000 Tilburg, LEPO. Box 90153, The Netherlands

large corpora combining with machine learning to analyze properties like multimodality and inferencing (Chen & Jhala, 2021; Iyyer et al., 2017). We here introduce one such corpus of researcher annotated comics with a focus on longitudinal and cross-cultural variance, the Visual Language Research Corpus (VLRC).

The VLRC differs from other corpora across two primary dimensions: the types of works analyzed (discussed below), and the type of annotations recorded. Many of the available corpora on comics have focused on annotation of fairly surface phenomena, driven by computer vision or basic theories of comics (Dunst et al., 2017; Fujimoto et al., 2016; Guérin et al., 2013). These features include annotation of panels, characters, balloons, captions, and text. In contrast, the VLRC contains annotations of comics across several structures specified within Visual Language Theory, a theoretical framework for graphic communication situated within the linguistic and cognitive sciences (Cohn, 2013).

Corpus analyses within the scope of Visual Language Theory were first undertaken outside of the VLRC, in focused projects that analyzed the framing structure of comic panels, particularly looking at how many characters appeared in each panel, the angle of viewpoint depicted in the frame, and the filmic shot scale for presentation of a panel's contents. These analyses also focused on comparisons between panels in comics from the United States and Japan (Cohn, 2011; Cohn et al., 2012), seeking empirical data to compare with McCloud's (1993) earlier claims that used no formalized methodology. These annotation efforts were precursors to the VLRC as early attempts to formalize analyses of Visual Language Theory at a time prior to the growth of linguistic and computational analyses of comics (for example, the original annotation of comics in Cohn, 2011 took place and were published in an online preprint in 2003), and thus reflect some of the earliest data-driven corpus analyses of comics.

Development of the VLRC itself began in the context of student projects at the University of California at San Diego with targeted topics decided by groups of student researchers. These projects often included varying substructures under analyses (framing, layout, multimodality, etc.) across different populations of comics (across cultures and time periods). Several of these studies resulted in publications exploring the way in which superhero comics from the United States changed across 80 years of publications in their page layouts (Pederson & Cohn, 2016) and their multimodal storytelling (Cohn et al., 2017), in comparisons of visual morphology between types of manga (Cohn & Ehly, 2016), and in cross-cultural comparisons of comics' layouts (Cohn et al., 2019), subjective viewpoint panels (Cohn et al., 2022), visual storytelling (Klomborg et al., 2022), and depictions of paths (Hacımusaoğlu & Cohn, 2022).

Subsequent annotation continued at Tilburg University undertaken as part of student theses. Comparisons were made across 80 years of published Dutch and Flemish comics both for their visual morphology (van Middelaar, 2017) and their storytelling (Dierick, 2017). Additional theses targeted the *Calvin and Hobbes* comic strip specifically, looking at how cartoonist Bill Watterson's style shifted over time related to page layouts (van Nierop, 2018) and storytelling (Schipper, 2018). Analyses of these aggregated annotations have yielded cross-cultural and longitudinal comparison of panel framing and layouts (Cohn, 2020) and of narrative patterns derived from the basic annotations of framing and situational changes between panels (Cohn, 2019).

Broadly, these studies aimed to further apply corpus methods to visual narratives using Visual Language Theory by analyzing a variety of different structures. An abiding question in this work was the range of cross-cultural diversity across comics of the world. Can identifiable patterns be observed between and within comics of different cultures? This question is essentially about the degree to which we can identify “visual languages” of the world. While analysis of patterns from different cultures provides a top-down method of quantitative comparison, substantive corpus analyses also allow statistical measures for visual languages to be identified bottom-up from similarities across the data (i.e., various types of clustering models). This cross-cultural focus led to analysis of a diverse sample of comics from across the United States, Asia, and Europe.

Within this focus on cross-cultural comparison, a related question asked about variation within a given culture’s comics. This question considers the consistency or diversity across comics from different genres or demographics, and about the degree to which there may exist multiple different visual languages within a culture, or if there are “varieties” in an overall shared visual language. Additional subtypes were thus analyzed in both comics from the United States and Japan. In U.S. comics, contrasts were made between “mainstream” comics, broadly consisting of superheroes and power fantasies, independent or “indie” comics, consisting of works outside the mainstream and typically published as “graphic novels”, and Original English Language (OEL) manga, which are comics imitative of Japanese manga but created by English speakers. Japanese manga were analyzed across four subtypes related to their stereotypes of the readership demographics: shonen manga (boys’), shojo manga (girls’), seinen manga (mens’), and josei manga (women’s).

An additional question investigated the degree to which visual languages within or across cultures have changed over time, which is a topic shared by other comics corpus projects (Bateman et al., 2021). This question pertained both to the observable changes that may occur for a given culture’s comics, but also whether influence across cultures’ comics might motivate such change. For example, the spread of manga from Japan throughout the world in the 1990s and early 2000s (Brienza, 2015) has resulted in observable influence of the “Japanese Visual Language” used in manga on other comics across the world. We therefore questioned whether this influence was quantitatively measurable, and whether imitators of manga (such as in the manga created by English speakers in the United States) indeed resembled the structures of comics from their intended style (manga) or from their cultural origin (the United States).

Altogether, these annotations have been compiled to comprise the full VLRC. The breadth of the types of information annotated across comics has yielded novel insights about their structure, while the breadth across comics from three continents provides important comparison of the diverse manifestation of visual narrative constructs. Below, we further describe the characteristics of the corpus itself, and describe its annotation fields.

2 Annotated documents

The VLRC consists of annotations for 376 stories from within 319 comic books and/or graphic novels from the United States, Asia, and Europe across several subgenres, as summarized in Table 1. These annotations amounted to 44,942 panels across 7,773 pages. As multiple coders sometimes annotated each comic, these numbers reflect the totals when averaging across coders and calculating the maximums annotated in each comic. When treating each annotation as unique and not collapsing across coders for stories, the corpus includes annotations of 78,805 panels across 11,413 pages in 491 comics.

As described above, works in the VLRC were initially based on fairly independent research projects and/or supplemented by convenience sampling. As a result, there

Table 1 Comics analyzed within the Visual Language Research Corpus, organized by global region

Country/Type	Years	# of stories	Total Pages	Total Panels	Panels/ Page
China	1946–2016	11	289	1,523	5.3
Japan (Josei)	2002–2015	10	225	1,091	4.8
Japan (Seinen)	1985–2014	10	210	1,043	5.0
Japan (Shojo)	1987–2010	22	694	3,481	5.0
Japan (Shonen)	2014–2014	54	1,383	7,140	5.2
Korea	2014–2014	15	316	1,327	4.2
Belgium/ Flanders	2002–2013	40	633	6,097	9.6
France	1981–2014	26	587	4,143	7.1
Germany	1987–2007	21	285	1,525	5.4
Netherlands	1940–2016	41	819	5,876	7.2
Spain	2012	1	24	105	4.4
Sweden	1980–2011	23	249	1,532	6.2
USA (Indie)	2002–2014	13	407	1,611	4.0
USA (Mainstream)	1940–2014	72	1,262	6,608	5.2
USA (US Manga)	1991–2006	17	390	1,840	4.7
Total		376	7,773	44,942	5.8
<i>Calvin and Hobbes</i>					
Daily strips	1985–1995	2,702		10,339	3.83
Sunday strips	1985–1995	453		4,389	9.69
Total		3,151		14,712	4.67

are inconsistent distributions of works from different cultures, but at least 10 comics were aimed to be analyzed from each country and/or subtype (e.g., demographics of Japanese manga), with the exception being the sole Spanish comic, which was the only one available at the time and then data collection ceased. In addition to this cross-cultural data, annotations were made of the complete run of the *Calvin and Hobbes* comic strip by Bill Watterson, totaling 14,712 panels across 3,151 strips.

Metadata with annotations included the title of each comic and its stories, along with any volume number, and title and volume number if they belonged to an anthology (if appropriate). We also include a listing of the artists, writers, and publisher of each comic and both the year of publication and original year of publication (if a reprint was annotated). Along with country and continent, comics were assigned to subtypes if they had them (e.g., demographics of Japanese manga or types of American comics), along with their genre (action, romance, etc.) and their format (comic book, comic strip, graphic novel, etc.).

Works in the VLRC were gathered via convenience sampling and/or from dedicated study of specific comics (e.g., analysis of *Calvin and Hobbes*). Many of the comics came from generous donations of physical books made by Antarctic Press, Archie Comics, Dark Horse Comics, Drawn & Quarterly, Fantagraphics Books, First Second Books, Humanoids Inc, IDW Publishing, NBM Publishing, NetComics, Oni Press, Top Cow, Top Shelf, Udon Entertainment, Vertical Inc, and Viz Media. We asked for no specific works in donations, and out of donated comics we attempted to choose works at random to reduce bias. Additional comics were gathered from public domain websites (www.comicbookplus.com) and/or purchasing comics to fit specific sampling criteria, such as to have five comics per decade in longitudinal analyses (of U.S. Mainstream comics, and comics from the Netherlands and Flanders).

Comics were independently hand-annotated using spreadsheets by 16 student researchers at the University of California at San Diego in the United States and at Tilburg University in the Netherlands. All students passed courses on Visual Language Theory (approximately 30 hours) and completed training and pre-annotation assessments in the schemes prior to annotating the corpus. 44% of the comic books (167/376) in the corpus were independently annotated by two researchers and were checked for interrater reliability. The VLRC includes the repeated analyses, distinguished by different annotations made for each annotator.

Each comic book was analyzed for the full book if it fell within an approximate target length (20–30 pages). If comics exceeded this length, researchers analyzed approximately the first 25 pages or 120 panels (rounded to the nearest page), whichever came first. Double-page spreads were given the number of the first page (i.e., pages 10 and 11 in a double-page spread would be entered as 10, skipping 11, with the next page being 12). Though we attempted to have rough parity in the number of panels per book and a minimum number of books per population, the variable length of books rendered inequalities in the overall quantities of panels per population.

Because the works within the VLRC are mostly protected by copyright and were made in ways dissociated from the original documents, often using physical copies of the books, only the annotations are shared. The full annotated corpus is stored in Excel and csv format along with documentation, and is openly available in both the

DataverseNL repository (Cohn, 2022, <https://doi.org/10.34894/LWMZ7G>) and from the Visual Language Lab website (<http://www.visuallanguage.com/vlrc>).

3 Annotation fields

The VLRC includes annotation of several different fields of information. However, not all annotation fields are applied consistently to all comics in the corpus due to the somewhat disparate projects with differing purposes that constituted the data in the VLRC. The distribution of annotation fields across the stories in the corpus is summarized in Table 2. For example, 20 Japanese manga were analyzed only for their morphology (Cohn & Ehly, 2016), while all stories in the corpus except those 20 manga were analyzed for their attentional framing, filmic shot scales, and semantic changes. Multimodality was only annotated for 62 Mainstream stories from the United States and in *Calvin and Hobbes* strips. *Calvin and Hobbes* strips were also annotated for idiosyncratic categories (such as whether Hobbes was shown as a real tiger or as a stuffed animal).

Below we describe the primary annotation fields within each of these categories, along with their definitions and criteria. Additional information about annotations can be found in the corpus documentation.

3.1 Attentional framing structure

Corpus studies of Visual Language Theory have consistently analyzed the way that panels frame their contents. Panels are made up of “active” information that contributes to the narrative sequence and “inactive” information that may provide meaning, but does not influence the sequential construal (Cohn, 2013). Because panels

Table 2 Total number of stories in the VLRC annotated for each of the annotation fields

Country	Attentional Framing	Filmic shot scales	Semantic Changes	Layout	Morphology	Path structure
China	11	11	11	9	-	7
Japan (Josei)	10	10	10	5	-	-
Japan (Seinen)	10	10	10	5	-	-
Japan (Shojo)	13	13	13	5	10	-
Japan (Shonen)	43	43	43	17	10	35
Korea	15	15	15	-	-	15
Belgium/Flanders	40	40	40	-	40	-
France	26	26	26	7	-	10
Germany	21	21	21	-	-	21
Netherlands	41	41	41	-	41	-
Spain	1	1	1	1	-	-
Sweden	23	23	23	23	-	-
USA (Indy)	13	13	13	12	-	-
USA (Mainstream)	66	66	66	58	40	15
USA (OEL Manga)	17	17	17	1	-	16

can vary in the amount of active information they depict, panels have been called “attention units”, which vary in their *attentional framing structure*, i.e., how a panel might window information in a scene to a reader. Variation in attentional framing can influence a sequence’s narrative structure (Cohn, 2015, 2019), and can modulate the processing of a visual sequence (Cohn & Foulsham, 2020; Foulsham & Cohn, 2021).

Attentional framing types for panels vary based on their amount of active or inactive information, summarized in the rows of Fig. 1. Panels depicting active entities include *macros* depicting multiple interacting active entities (e.g., typically characters), *monos* showing only single entities, and *micros* framing less than a single entity, typically with a close up on a portion of the face or any other body part. *Amorphic* panels show no active entities, depicting only environmental information.

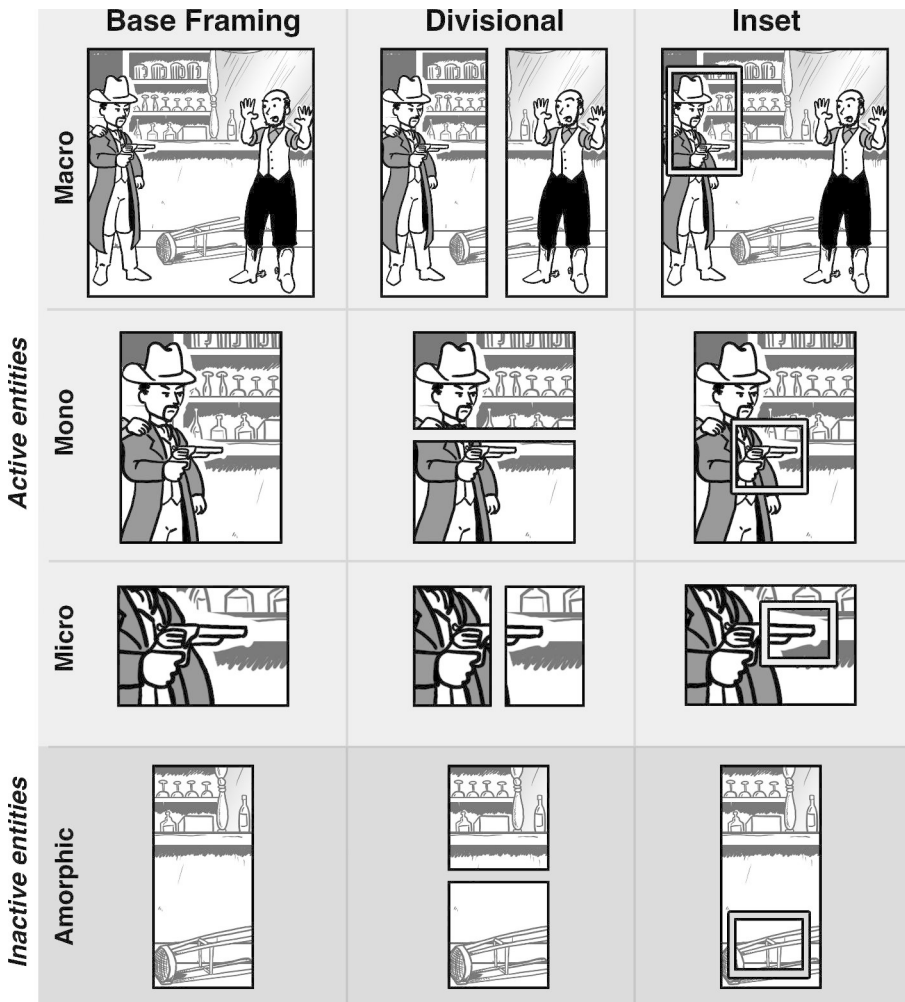


Fig. 1 Attentional framing types modulating the number of active entities in a panel, and modifications of these types through additional framing devices

Panels that did not clearly fall into these categories were deemed as ambiguous (e.g., such as black panels, or those with only text).

Framing can be further modified by manipulating the *paneling structure*, how panel framing relates to the contents of other panels, as in the columns of Fig. 1. A *Base* representation of a panel is the framing applied when a panel stands on its own. A *Divisional* panel belongs to two or more panels which together depict a larger image. An *Inset* is a panel that is encapsulated by a surrounding, *Dominant* panel.

3.2 Filmic shot scale

Along with attentional framing structure, panels can also modify the presentation of their content, similar to the way that films structure their shots. Filmic shot scales borrow the classifications from film theory (Bordwell & Thompson, 1997; Cutting, 2015). Long shots contained a predominant view of the scene including the figures inside it. Full shots included a figure's whole body, while medium shots included a figure's waist and above. Medium close shots included above a figure's bust, while close shots showed the head and/or shoulders. Extreme close ups showed aspects of face and/or other zoomed-in body parts. An ambiguous category was applied for panels without clear shots scales, such as black panels or those with only text.

3.3 Semantics of and between panels

An additional primary area of analysis was how meaning changed across each juxtaposed panel in a comic. While McCloud's (1993) analysis of the "transitions" between panels gave way to several inventories of how meaning changes across sequential images (Bateman & Wildfeuer, 2014; Gavalier & Beavers, 2018; Saraceni, 2016; Stainbrook, 2016), psychological theories of visual narrative comprehension have argued that readers track incremental changes across multiple situational dimensions at once (Cohn, 2020; Loschky et al., 2020). In the VLRC, annotations were made across three primary situational dimensions informed by narrative and discourse research (Magliano & Zacks, 2011; Zwaan & Radvansky, 1998): changes between *characters*, between *spatial locations*, and between *states of time*. We considered situational changes to be non-mutually exclusive and non-exhaustive, meaning that multiple changes could occur at the same time and changes could be both full and partial.

Our annotation protocols followed those used to analyze previous experimental stimuli (Cohn & Bender, 2017). Full shifts in characters between panels were given a "1", partial changes were given a ".5" (i.e., one or more characters were added or omitted while other stayed the same), and no change was given a "0". Full shifts in spatial location were given a "1", partial changes (such as shifting within a common space, like between rooms in the same building) a ".5", and no changes, maintaining the same location, a "0." If time was interpreted as passing between panels relations were given a "1", but no such inference of time passing could be made, or it was ambiguous between panels, it was given a "0". An example analysis for a comic page from *The Amazing Spider-Man* from within the VLRC is provided in Fig. 2.

Time change	Character change	Spatial location change
Yes	Full	Full
Yes	Partial	None
No	Full	Partial

Fig. 2 Example annotation of the situational changes between panels across dimensions of time, characters, and spatial location for a page within the VLRC. All analyses were made for a panel relative to its previous panel, and thus arrows are depicted going backwards. *The Amazing Spider-Man #539* by J. Michael Straczynski and Ron Garney. Spider-Man © Marvel Comics

An additional situational characteristic of panels themselves was analyzed, with coding of whether a panel used a subjective viewpoint, i.e., a “point-of-view” shot where the panel depicted the viewpoint of a character in the scene.

3.4 Layout (external compositional structure)

The physical arrangements of page properties were studied as part of their *external compositional structure*, or layout. The VLRC analysis of layout focused on a “bottom-up” approach to characterizing the specific properties and arrangements of panels within each page. This differs from “top-down” approaches which characterize each whole page with a single classification (e.g., Bateman et al., 2021; Bateman et al., 2016; Wildfeuer et al., 2022).

Analyses of layout were made across several dimensions including panel properties, proximity between panels, panel arrangements, and directionality. Panel properties were based on the features of panels themselves. Panel shapes included more standard shapes like squares or rectangles, but also less typical shapes like circles, triangles, irregular shapes (without any distinct geometry), and diagonals (as if spanning from opposite corners of a square). The presence or absence of panel borders were also noted. *Borderless panels* were images with no depicted frame around them, while *bleeding panels* (Fig. 3i) were borderless panels where any side extended beyond the edge of the page boundary.

A first dimension of panels' relations was their relative proximity. This was primarily assessed by the size of the “gutter”, the physical space between panels. *Normal gutters* were determined as a standard width between panels, based on the common practices of a given book. A *separation* (Fig. 3g) was a gutter that extended beyond this “standard” distance. When gutters were nonexistent, such as with only a line drawn between panels, they were annotated as having *no gutter*, while an *overlap* (Fig. 3h) was where a panel was placed into the space of another panel. An *inset* (Fig. 3f) was a panel placed entirely inside of another, *dominant* (Fig. 3e) panel.

Panel arrangements related to patterned ways that panels were organized relative to each other, the most basic of which was a *grid-type layout*, where panels occupied rows stacked vertically. A *pure grid* (Fig. 3a) maintained contiguity between both horizontal and vertical borders of juxtaposed panels to create a “+” shaped junction between panels, while *vertical staggering* (Fig. 3b) and *horizontal staggering* (Fig. 3c) had panel borders that were not contiguous within an otherwise grid-like layout (i.e., retaining only one contiguous gutter). In *blockage* (Fig. 3d), a vertical column was placed next to a larger, longer panel to create a “T” shaped junction between panels. A *whole row* (Fig. 3e) panel extended the full width of a page, while a *whole column* panel extended the length (top-to-bottom) of a page. A *whole page* was a “splash page” where one panel occupied the entire page.

Finally, we assessed the directionality between panels: the orientation of panels relative to each other, as defined by approximating the centerpoint of a panel in relation to the centerpoint of the narratively preceding panel. The vector formed between these points was annotated as one of eight directionalities (*right, left, up, down*, and angular directions in-between these categories). A 2×2 grid would thus have directions of right, down-left, right. This would render the first page of Fig. 3 with direc-

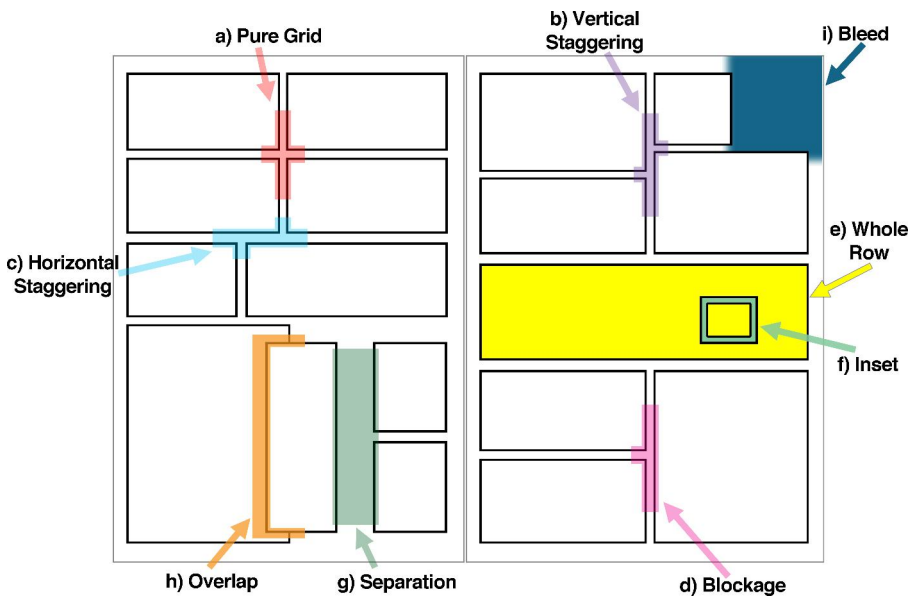


Fig. 3 Examples of panel arrangements analyzed within the VLRC

tions of right, down-left, right, down-left, right, down-left, right, up-right, and down, while the second page of that figure would be down, up-right, down, down-left, right, down-left, down, and up-right. The starting panel of a page was recorded as being the *first panel*, with no directionality, because it had no preceding panel.

3.5 Multimodality

Multimodal relationships were originally analyzed following the framework in Cohn (2016b) and its decision tree for categorical assignments. However, we have altered the terminology within the VLRC to adapt to changes in the framework introduced in Cohn and Schilperoord (2022) which provided more nuance to the two dimensions of multimodality which were concatenated in the categories of Cohn (2016b). These dimensions were: how each modality contributed to the overall meaningful gist (*semantic weight*), and what structural features were displayed by each modality (*grammatical symmetry*).

First, annotators assessed the semantic weight of the multimodal interaction, defined as the relative contribution of modalities to the overall gist. Semantic weight was decided by whether the overall gist of a panel was retained when imagining each modality as omitted. If the gist stayed apparent during omission, the retained modality was inferred as more semantically weighted (Visual-weight, Verbal-weight), but if the gist was not retained, meaning was inferred to be shared between modalities (Balanced-weight).

Annotators next assessed the relative contributions of grammatical structures for each modality. In this context, panels were assessed for which structures they demonstrated, whether it contained text with a syntactic structure versus those with single units (like onomatopoeia), or whether the panel was placed in a sequence that used a coherent narrative structure, as assessed using diagnostic tests (Cohn, 2015). Here, annotators decided whether both modalities used a complex grammatical structure (syntax, narrative), which was deemed *symmetrical* (previously “assertive”) if both were present. If only one modality used a complex grammar and the other used a simple grammar (such as a single word), they were assigned as using an *asymmetrical* interaction (previously “dominant”). The direction of asymmetry was specified through the semantic weight (i.e., the more weighted modality was also the more structurally complex one). Panels with only one modality were deemed *unimodal* (previously “autonomous”).

To further analyze whether these multimodal interactions were affected by the quantity of words, annotators also counted the total number of words per panel.

3.6 Path structure

We additionally analyzed the way that motion events were depicted in comic panels, particularly for their paths. Paths are the traverse taken by an object in motion, characterizing it going *from* a place *to* another place. This analysis was undertaken because of observations across linguistic typologies in differences with how spoken languages structure motion events (Talmy, 2000), and indeed path structure within

the analyzed comics appeared to reflect some typological properties of the authors' spoken languages (Hacımusaoğlu & Cohn, 2022).

Annotations counted the number of times a given panel represented a path by depicting its *source* (starting point), *route* (midpoint and the path itself), and/or *goal* (endpoint). Cues used to signal these paths were simultaneously annotated, be they graphic devices like *motion lines* (lines attached to a moving figure), *suppletion* (lines replacing the whole or part of a figure), *polymorphism* (repeated figures), *backfixes* (lines set behind a moving figure), or the *postural cues* of figures in motion.

3.7 Visual morphology

We analyzed the closed-class morphology (Cohn, 2013) or “symbology” for several comics in the corpus. This effort had begun with analysis of 73 visual morphemes compared in shonen and shojo manga (Cohn & Ehly, 2016) which were then expanded to include 156 total morphemes analyzed in Dutch and Flemish comics (van Middelaar, 2017). As this list of morphemes is fairly extensive, readers are directed for further details to these papers, and to the data dictionary for the VLRC.

Along with various idiosyncratic visual morphemes, we included several morphological classes with various individual types (Cohn, 2013). *Upfixes* were elements that floated above characters' heads (24 morphemes, ex. hearts, stars, gears, exclamation marks, question marks), and *eye-umlauts* were symbols that replaced characters' eyes (7 morphemes, ex. hearts, spirals, stars, etc., plus 8 additional manga eye variations). Backgrounds were analyzed if they carried specific symbolic meaning (8 morphemes), such as representing weather for emotional purposes (a storm as anger or depression), blackness (for moodiness), or with lines set behind a moving figure (backfixing lines), among others.

Finally, carriers were holders of text, conventionally recognized as speech balloons, thought bubbles, captions, and sound effects. However, Visual Language Theory generalizes “carriers” across these visual representations, and uses the categorization of public, private, non-sentient, or satellite carriers based on the degree to which characters in the scene had access to the content of these carriers (Cohn, 2013). Surface representations of carriers were then annotated for whether they were shown with a defined border, with a tail, or for specific representations (i.e., as balloons, bubbles, captions, etc.). Carriers were analyzed alongside other morphemes in Dutch and Flemish comics, but were the only morphology analyzed in a subsection of Mainstream comics from the United States.

4 Advantages and limitations

Analysis of the VLRC thus far has yielded insights into the way that comics differ across cultures and change across time. It has provided important data for starting to understand the dimensions of cross-cultural diversity across cultures' comics and has laid the groundwork for assessing different “visual languages” spanning within and across cultures. Such research has both compared the data between countries

and subtypes, and has used statistical clustering to assess how structures used within books might transcend cultures to characterize broader varieties (Cohn, 2020).

In this regard, the VLRC is a foundational dataset for asking further questions about the structure of visual narratives by uniquely annotating comics with theoretical constructs from Visual Language Theory, rather than physical features of page composition (Fujimoto et al., 2016; Guérin et al., 2013) or surface aspects of comics content (Dunst et al., 2017). As such, the VLRC is part of the broader Visual Language Theory research program to understand the cognitive representations involved in understanding visual narratives by combining theory with corpus analysis and experimentation (Cohn, 2016a).

To this end, analysis of the VLRC thus far has already informed experimentation. For example, analysis of narrative constructions in the VLRC found a pattern that was more frequent in Japanese manga than American and European comics (Cohn, 2019), and indeed brainwaves evoked by this pattern were modulated by frequency of readership in manga (Cohn & Kutas, 2017). These studies thus show how comprehension is modulated by distinct cultural patterns and highlight the ways that corpus analyses can inform the design and interpretation of psychological experimentation.

Despite these insights, it is worth also highlighting the limitations of the VLRC as a corpus. Because of the varied projects that contributed to the corpus, there is heterogenous coverage of annotation fields across the works within the corpus. In addition, there is inconsistencies for the corpus's aim of cultural diversity, focusing on comics from the United States, northwestern Europe, and East Asia. While these regions are representative of prominent areas with thriving comics industries, a truly cross-cultural corpus should also include comics with a complete global scope. This disparate nature of the VLRC data and constrained cross-cultural coverage make direct comparison of annotation fields across the whole corpus more limited and the scope of cross-cultural inferences more constrained, despite the insightful analyses that remain across the varied works that are present.

An additional limitation arises from the methods used to annotate the comics in the corpus. Despite reflecting a visual and multimodal medium, we here were not able to annotate the visual properties of these media directly, as data in the VLRC was not gathered using graphically-based annotation software (Dunst et al., 2017), but rather analyzed "by hand" using spreadsheets, often while looking at physical copies of comics. This conversion of analog visual representations to a simple digital spreadsheet format resulted in needing to choose our basic unit of annotation, which here was the "panel." While category assignment could readily be made for whole panels easily, this meant that all information smaller than a panel (characters, visual morphology) were tabulated numerically within each panel, which lacks precision. In addition, some aspects of the spatial configurations of layouts may be lost or complicated given the binary categorical designations given to panels as the unit of analysis.

A related limitation was thus that annotations included no features of the physical properties of the documents themselves, such as sizes of or angles between areas of interest. The VLRC dataset on its own is thus not suitable for analyses using computer vision (Fujimoto et al., 2016; Guérin et al., 2013), although bibliographic information is provided should researchers want to pair VLRC data with the original works independently.

A solution to these limitations is underway in current work, where we have developed a Multimodal Annotation Software Tool (MAST) to facilitate further visual and multimodal annotation (Cardoso & Cohn, 2022). This tool allows for the selection of regions along a visual surface that can then be annotated with any prespecified scheme, and relations can also be established between annotations to create dependencies and hierarchies. Because MAST facilitates the recording of visual regions, more precision is possible for assessing spatial dimensions like layouts, while also providing data about regions' spatial arrangements, their relative area or size, and/or their angles from each other. This information is stored in a server that facilitates collaborative annotation efforts and enforces varied permissions for accessing data and annotated documents. Our current efforts aim to use MAST to construct a corpus with a global scope with more consistent coverage of annotations. We therefore seek to overcome the limitations in the VLRC while further investigating the types of questions that it first addressed, and to also provide resources for other researchers.

Altogether, the VLRC provides a valuable starting corpus for the analysis of cross-cultural diversity in the structures of comics, while interfacing with the constructs from linguistic and cognitive research on those materials. It thus provides a rich source of material to be analyzed in its own right, and lays the groundwork for future research.

Acknowledgements The VLRC annotations were made by student researchers who include: Jessika Axné, Justin Brookshier, Michaela Diercks, Mark Dierick, Sean Ehly, Ryan Huffman, Kaitlin Pederson, Ryan Taylor, Lincy van Middelaar, Vivian Wong, Rebecca Yeh.

Funding Statement This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 850975).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bateman, J. A., Veloso, F. O. D., & Lau, Y. L. (2021). On the track of visual style: A diachronic study of page composition in comics and its functional motivation. *Visual Communication*, 20(2), 209–247. <https://doi.org/10.1177/1470357219839101>.
- Bateman, J. A., Veloso, F. O. D., Wildfeuer, J., Cheung, F. H., & Guo, N. S. (2016). An Open Multilevel classification Scheme for the Visual Layout of Comics and Graphic Novels: Motivation and design. *Digital Scholarship in the Humanities*, 32(3), 476–510. <https://doi.org/10.1093/llc/fqw024>.
- Bateman, J. A., & Wildfeuer, J. (2014). Defining units of analysis for the systematic analysis of comics: A discourse-based approach. *Studies in Comics*, 5(2), 373–403. https://doi.org/10.1386/stic.5.2.373_1
- Bordwell, D., & Thompson, K. (1997). *Film Art: An Introduction* (5th Edition ed.). New York, NY: McGraw-Hill.

- Brienza, C. (2015). *Global Manga: "Japanese" Comics without Japan?* Surrey, UK: Ashgate Publishing, Ltd.
- Cardoso, B., & Cohn, N. (2022). The Multimodal Annotation Software Tool (MAST). In Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022) (pp. 6822–6828). Marseille, France: European Language Resources Association.
- Chen, Y. C., & Jhala, A. (2021). A Computational Model of Comprehension in Manga Style Visual Narratives. Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43).
- Cohn, N. (2011). A different kind of cultural frame: An analysis of panels in american comics and japanese manga. *Image [&] Narrative*, 12(1), 120–134.
- Cohn, N. (2013). *The visual language of comics: Introduction to the structure and cognition of sequential images*. London, UK: Bloomsbury.
- Cohn, N. (2015). Narrative conjunction's junction function: The interface of narrative grammar and semantics in sequential images. *Journal of Pragmatics*, 88, 105–132. <https://doi.org/10.1016/j.pragma.2015.09.001>.
- Cohn, N. (2016a). Interdisciplinary approaches to visual narrative. In N. Cohn (Ed.), *The visual narrative reader* (pp. 1–15). London, UK: Bloomsbury.
- Cohn, N. (2016b). A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition*, 146, 304–323. <https://doi.org/10.1016/j.cognition.2015.10.007>.
- Cohn, N. (2019). Structural complexity in visual narratives: Theory, brains, and cross-cultural diversity. In M. Grishakova, & M. Poulaki (Eds.), *Narrative complexity: Cognition, Embodiment, Evolution* (pp. 174–199). Lincoln: University of Nebraska Press.
- Cohn, N. (2020). *Who understands comics? Questioning the universality of visual language comprehension*. London: Bloomsbury.
- Cohn, N. (2022). The Visual Language Research Corpus (VLRC) Project. doi:<https://doi.org/10.34894/LWMZ7G>
- Cohn, N., Axnéer, J., Diercks, M., Yeh, R., & Pederson, K. (2019). The cultural pages of comics: Cross-cultural variation in page layouts. *Journal of Graphic Novels and Comics*, 10(1), 67–86. <https://doi.org/10.1080/21504857.2017.1413667>.
- Cohn, N., & Bender, P. (2017). Drawing the line between constituent structure and coherence relations in visual narratives. *Journal of Experimental Psychology: Learning Memory & Cognition*, 43(2), 289–301. <https://doi.org/10.1037/xlm0000290>.
- Cohn, N., & Ehly, S. (2016). The vocabulary of manga: Visual morphology in dialects of Japanese Visual Language. *Journal of Pragmatics*, 92, 17–29. <https://doi.org/10.1016/j.pragma.2015.11.008>.
- Cohn, N., & Foulsham, T. (2020). Zooming in on the cognitive neuroscience of visual narrative. *Brain and Cognition*, 146, 105634. <https://doi.org/10.1016/j.bandc.2020.105634>.
- Cohn, N., Hacimusaoğlu, I., & Klomberg, B. (2022). The framing of subjectivity: Point-of-view in a cross-cultural analysis of comics. *Journal of Graphic Novels and Comics*, 1–15. <https://doi.org/10.1080/21504857.2022.2152067>.
- Cohn, N., & Kutas, M. (2017). What's your neural function, visual narrative conjunction? Grammar, meaning, and fluency in sequential image processing. *Cognitive Research: Principles and Implications*, 2(27), 1–13. <https://doi.org/10.1186/s41235-017-0064-5>.
- Cohn, N., & Schilperoord, J. (2022). Remarks on multimodality: Grammatical interactions in the parallel architecture. *Frontiers in Artificial Intelligence*, 4, 1–21. <https://doi.org/10.3389/frai.2021.778060>.
- Cohn, N., Taylor, R., & Pederson, K. (2017). A picture is worth more words over time: Multimodality and narrative structure across eight decades of american superhero comics. *Multimodal Communication*, 6(1), 19–37. <https://doi.org/10.1515/mc-2017-0003>.
- Cohn, N., Taylor-Weiner, A., & Grossman, S. (2012). Framing attention in Japanese and American Comics: Cross-cultural differences in attentional structure. *Frontiers in Psychology - Cultural Psychology*, 3, 1–12. <https://doi.org/10.3389/fpsyg.2012.00349>.
- Cutting, J. E. (2015). The framing of characters in popular movies. *Art & Perception*, 3(2), 191–212. <https://doi.org/10.1163/22134913-00002031>.
- Dierick, M. (2017). Visual storytelling in Dutch and Flemish comic books: A corpus analysis across culture and time. (Bachelors Thesis). Tilburg University, Tilburg, Netherlands.
- Dunst, A., Hartel, R., & Laubrock, J. (2017). 9–15 Nov. 2017). The Graphic Narrative Corpus (GNC): Design, Annotation, and Analysis for the Digital Humanities. Paper presented at the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR).
- Foulsham, T., & Cohn, N. (2021). Zooming in on visual narrative comprehension. *Memory & Cognition*, 49(3), 451–466. <https://doi.org/10.3758/s13421-020-01101-w>.

- Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T., & Aizawa, K. (2016). Manga109 dataset and creation of metadata. Paper presented at the Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding, Cancun, Mexico. <https://doi.org/10.1145/3011549.3011551>
- Gavaler, C., & Beavers, L. A. (2018). Clarifying closure. *Journal of Graphic Novels and Comics*, 11, 182–211. <https://doi.org/10.1080/21504857.2018.1540441>
- Guérin, C., Rigaud, C., Mercier, A., Ammar-Boudjelal, F., Bertet, K., Bouju, A., & Revel, A. (2013). eBDtheque: a representative database of comics. Paper presented at the Document Analysis and Recognition (ICDAR), 2013 12th International Conference on.
- Hacimusaoğlu, I., & Cohn, N. (2022). Linguistic typology of motion events in visual narratives. *Cognitive Semiotics*, 1–26. <https://doi.org/10.1515/cogsem-2022-2013>.
- Iyyer, M., Manjunatha, V., Guha, A., Vyas, Y., Boyd-Graber, J., Daumé, H. III, & Davis, L. (2017). *The amazing Mysteries of the gutter: Drawing Inferences between Panels in Comic Book Narratives* (pp. 6478–6487). CVPR.
- Klomborg, B., Hacimusaoğlu, I., & Cohn, N. (2022). Running through the who, where, and when: A cross-cultural analysis of situational changes in Comics. *Discourse Processes*, 59(9), 669–684. <https://doi.org/10.1080/0163853X.2022.2106402>.
- Laubrock, J., & Dunst, A. (2020). Computational approaches to comics. *Topics in Cognitive Science*, 12(1), 274–310. <https://doi.org/10.1111/tops.12476>.
- Loschky, L. C., Magliano, J., Larson, A. M., & Smith, T. J. (2020). The Scene Perception & Event Comprehension Theory (SPECT) Applied to visual narratives. *Topics in Cognitive Science*, 12(1), 311–351. <https://doi.org/10.1111/tops.12455>.
- Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, 35(8), 1489–1517. <https://doi.org/10.1111/j.1551-6709.2011.01202.x>.
- McCloud, S. (1993). *Understanding Comics: The invisible art*. New York, NY: Harper Collins.
- Neff, W. A. (1977). *The Pictorial and Linguistic Features of Comic Book Formulas*. (Doctoral Dissertation). University of Denver, Denver, CO.
- Pederson, K., & Cohn, N. (2016). The changing pages of comics: Page layouts across eight decades of american superhero comics. *Studies in Comics*, 7(1), 7–28. https://doi.org/10.1386/stic.7.1.7_1.
- Saraceni, M. (2016). Relatedness: Aspects of textual connectivity in comics. In N. Cohn (Ed.), *The Visual Narrative Reader* (pp. 115–129). London: Bloomsbury.
- Schipper, D. (2018). *Visual storytelling in Calvin and Hobbes*. (Bachelors Thesis). Tilburg University, Tilburg, Netherlands.
- Stainbrook, E. J. (2016). A Little Cohesion between Friends; Or, We're Just Exploring Our Textuality: Reconciling Cohesion in Written Language and Visual Language. In N. Cohn (Ed.), *The Visual Narrative Reader* (pp. 129–154). London: Bloomsbury.
- Talmy, L. (2000). *Toward a cognitive semantics* (2 vol.). Cambridge, MA: MIT Press.
- van Middelaar, L. (2017). *It ain't much, if it ain't Dutch: Visual morphology across eight decades of Dutch and Flemish comics*. (Bachelors Thesis). Tilburg University, Tilburg, Netherlands.
- van Nierop, J. (2018). *Scientific Progress Goes 'Boink': A systematic study in page layout of Calvin and Hobbes*. (Bachelors Thesis). Tilburg University, Tilburg, Netherlands.
- Wildfeuer, J., van der Sluis, I., Redeker, G., & van der Velden, N. (2022). No laughing matter!?! Analyzing the page layout of instruction Comics. *Journal of Graphic Novels and Comics*, 1–22. <https://doi.org/10.1080/21504857.2022.2053559>.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.