Using ATLAS.ti for constructing and analysing multimodal social media corpora

Yuze Sha[a] and Beth Malory[b*]

[a]Lancaster University

[b]University College London

*b.malory@ucl.ac.uk

Abstract

Methods to enable comprehensive corpus analyses of multimodal data are imperative to furthering our understanding of social media and digital communication. In this study, we demonstrate how ATLAS.ti version 24.2.0 can be used to construct such corpora and conduct multimodal corpus-assisted discourse studies (MCADS). The focus for such studies may be the exploration of complex patterns of co-occurrence, both intra- and inter-mode, or iterative corpus queries, especially when unexpected patterns move the research focus beyond initial research questions. In such cases, ATLAS.ti's functionalities facilitate a triangulation of automatic pattern recognition and in-depth manual analysis. It supports a flexible, user-defined approach to the multimodal analysis of short-form social media datasets, overcoming traditional limitations such as analyses being restricted to emojis, or pre-set thematic dimensions in current AI software. In this way, the proposed methodology enables in-depth MCADS, as illustrated in the case study presented in this paper, which analyses the co-occurrences of evaluations and visual representations of social actors.

Keywords: corpus linguistics, multimodal discourse, multimodal corpus construction, social media discourse, language-image relations

1. Introduction

This paper introduces a novel methodology for constructing, annotating, and analysing multimodal corpora in social media discourse using ATLAS.ti 24.2.0. Originally designed for qualitative data analysis, we repurpose ATLAS.ti to demonstrate its effectiveness in corpus-based analysis of multimodal short-form social media discourses. Its ability to map and visualise co-occurrences within and across modes, as well as its flexible query tools for multimodal datasets, distinguish it from traditional corpus analysis tools.

The following sections begin with a review of previous studies in multimodal corpus construction and analysis. We then provide a detailed overview of ATLAS.ti, outlining its application in building a multimodal corpus of social media data. Our methodology centres on four analytical functions: frequency analysis, co-occurrence analysis within and across modes (and its visualisation with the Force-Directed Graph capability), query tools, and multimodal concordance analysis. This helps us first identify quantitative patterns of features and variants in the corpus, then interpret them qualitatively, accounting for contextual factors (Baker and Egbert 2016).

We use a case study on the multimodal representation of #freerange on Twitter (Malory and Sha) to showcase this approach. In presenting this case study, we aim to address common challenges in multimodal corpus analysis of social media discourses, showcasing how ATLAS.ti can be a powerful tool for researchers in this field.

2. Corpus Linguistics & Multimodal Social Media Discourse

Traditionally, corpus studies have focused primarily on monomodal linguistic datasets (e.g., Baker et al. 2008; Baker and McEnery 2005 and 2015; Gillings et al. 2023). However, given that communication is inherently multimodal, restricting research to language-only corpora overlooks the distinctive contribution of each semiotic mode involved, as well as interactions between modes. This oversight is highlighted by Caple (2018: 85) as a significant "blind spot" in corpus approaches to discourse studies.

In recent times, there has been a discernible trend towards attempting to integrate quantitative methods into multimodal discourse studies. Allwood (2008: 208) defines a multimodal corpus as a collection of language and communication materials accessible through computer-aided tools, encompassing various modes. Methodological experimentation in this field has expanded, ranging from manual to fully automated computerised analyses. The present study will focus on the multimodal construction and analysis of short-form social media corpora, which has previously been the focus of only a few studies (e.g., Christiansen et al. 2020; Collins 2020).

2.1. Multimodal corpus studies of social media discourse

Social media posts are inherently multimodal, combining semiotic resources such as emojis, memes, videos, and external links. Their diverse modes, shaped by platform-specific properties, have increasingly become the focus of many multimodal discourse studies in recent years (Bouvier and Machin 2020; Djonov and Zhao 2013). For example, Han and Zappavigna (2024) examined TikTok videos, conducting a qualitative analysis of rhythm across speech, bodily action, gesture, and music. Andersson (2024) employed a qualitative, social semiotic discourse analysis of YouTube reaction videos on transgender activism, demonstrating how the

integration of different modes can express impoliteness towards non-targeted viewers. Both studies focus on video-based platforms; however, for social networking platforms such as Twitter and Instagram, most social media discourse studies emphasise images and written language in posts (Jovanovic and van Leeuwen 2018).

While research on social media discourse has grown, it remains largely divided into two methodological camps: qualitative (e.g., Chałupnik and Brookes 2022; Hansson and Page 2023; Jovanovic and van Leeuwen 2018; Ng 2018) and quantitative (e.g., Christiansen et al. 2020; Collins 2020) approaches, each with inherent limitations.

Quantitative/statistical methods, particularly those using computational tools like Google Cloud Vision (e.g., Christiansen et al. 2020), enhance efficiency when handling large datasets. However, these automated tools sometimes struggle to capture the nuance and granularity of visual expressions (Baker and Collins 2023). Also, though valuable for identifying large-scale patterns, purely quantitative approaches often overlook the intricate semantic and functional relationships across modes.

In contrast, qualitative discourse studies with human interpretation (e.g., Chałupnik and Brookes 2022; Ng 2018), as Christiansen et al. (2020) assert, remain crucial for understanding the complexities of multimodal social media content, especially uncovering discourses not explicitly represented in language. However, purely qualitative multimodal discourse studies also present challenges. For instance, Widdowson (2000) highlighted the potential for "cherry-picking", where qualitative analysts selectively focus on atypical data to support preconceived arguments. Additionally, the manual analysis process often struggles to reveal cross-modal relationships across posts, limiting researchers' ability to objectively identify prominent patterns.

Consequently, there is a clear need for research which combines the benefits of quantitative and qualitative methods for multimodal discourse analysis on social media. Quantitative methods can enhance pattern recognition and mitigate the risks of selective interpretation (Baker and Levon 2015). Such methods can reveal patterns in language use, especially co-occurrences, that may not be immediately evident to human intuition (Widdowson 2000). Meanwhile, qualitative methods allow exploration of the nuanced relationships across modes and facilitate interpretation of patterns within broader social, political, and cultural contexts in a way that is not possible using quantitative methods (Baker et al. 2008). Together, these approaches can offer a more comprehensive framework for understanding multimodal communication on social media platforms.

Efforts within corpus linguistics to study multimodal social media discourse are relatively new, and, to our knowledge, no corpus software has the capability yet to fully interpret multimodal data in the same way it handles linguistic monomodal data, for example identifying co-occurring patterns and demonstrating associated concordances across different modes. Researchers have attempted to bridge this gap by adapting existing corpus tools, though these adaptations are mostly applicable to specific target semiotic sources, such as emojis (e.g., Zappavigna and Lorenzo 2024), which are readable in some already existing linguistic corpus software (e.g., Sketch Engine). Some studies have encoded emojis with Unicode values to make them more machine-readable; for instance, Collins (2020) focused on emojis and images, by converting emojis in Facebook posts to their Unicode values and categorising images into predefined groups, enabling searches comparable to linguistic queries. Other approaches attempting to take account of multimodality, such as Bednarek and Caple's (2017) research, combined results from separate corpus analysis of linguistic data and qualitative analysis of sampled visual data. The brevity of much social media data presents an additional challenge

for traditional corpus software, since fragmented meaning-making units and contextual interpretations (Clarke 2022) necessitate a different approach than would be taken with traditional media discourses. This highlights the ongoing need for more advanced, integrated analysis tools and approaches capable of addressing the intricacies of multimodal social media data.

2.2. Introducing ATLAS.ti for Multimodal Corpus Studies of Social Media Discourse

To address the existing gaps in multimodal corpus studies of social media discourse, this paper introduces the application of ATLAS.ti 24.2.0, an interpretive piece of software originally developed for computer-assisted qualitative data analysis (CAQDAS) in 1993. Aligned with grounded theory principles, CAQDAS tools, such as ATLAS.ti, allow the iterative development of data categories (Page 2022), enabling more transparent, systematic, and replicable analytical processes (Woods et al. 2016).

ATLAS.ti 24.2.0 is available in both Desktop (offline) and Web (online) versions, with a feature comparison provided on the official ATLAS.ti website (2024a). Researchers can access a free trial of the software, though continued use typically requires a paid license. The software includes an inter-coder agreement mode to support collaboration among multiple users. Additionally, a technical document detailing ATLAS.ti's concepts and functionalities is available to assist researchers with limited prior knowledge (see ATLAS.ti GmbH 2024b).

ATLAS.ti has a strong track record of application in linguistic monomodal discourse and conversation studies (e.g., Friese et al. 2018; O'Halloran 2011; Potgieter 2014; Karera et al. 2024). Its reach also extends beyond linguistics to fields such as business strategy analysis; for example, Karera et al. (2024) used the software to conduct content analysis of semi-structured interviews with medical officers, and Potgieter (2014) applied it to summarise customer interaction strategies on social media in classroom settings. One of ATLAS.ti's key advantages is its ability to incorporate multimodal datasets (Antoniadou 2017), which has led to its use in a limited number of qualitative multimodal discourse studies. For instance, Rossolatos (2014) employed ATLAS.ti 7 to analyse the multimodal rhetorical structure of TV advertisements. The researcher identified their minimal units of analysis (i.e., "ad filmic segments") by focusing on five groups of pro-filmic elements, such as actors, setting changes, and slogans, which were then coded for rhetorical figures. However, due to the comparatively limited capacity of ATLAS.ti 7 at that time, the summary only employed descriptive statistics (e.g., frequency), and the segmentation process employed poses issues, due to the overlap of some coded elements, for example setting changes and slogans, in the timeline.

In contrast, ATLAS.ti 24.2.0 is particularly well-suited for analysis of data from text-and-image-based social media platforms. The short-form nature of posts, often limited by character constraints, allows researchers to treat each post as a discrete analytical unit. This enables ATLAS.ti's functionalities to be applied in constructing and analysing multimodal social media texts, much like in monomodal corpus studies. Researchers can use these functionalities to identify both intra- and inter-mode co-occurrences, which can then be systematically traced, visualised, and analysed both quantitatively and qualitatively (as detailed in Sections 3 and 4).

Despite its potential, the use of ATLAS.ti for short-form multimodal social media discourse studies is still in its infancy. Most studies have focused on monomodal data, such as the linguistic analysis of blogs (Paulus and Lester 2014), or traditional content analysis methods like theme identification in tweets about parks during COVID-19 (Choe et al. 2022). This

highlights the need for a more sophisticated, multimodal analytical framework for social media discourse analysis with ATLAS.ti.

ATLAS.ti 24.2.0 has introduced new features such as concept cloud generation, opinion mining, sentiment analysis, and AI-based summaries. While these tools offer valuable insights, especially for large-scale pattern recognition, our pilot study revealed that they may not yet be robust enough to fully replace human interpretation or advanced NLP methods. For this reason, the present study does not rely on these automatic features but instead emphasises the interpretive potential of ATLAS.ti in combination with manual analysis.

This paper aims to demonstrate how ATLAS.ti can be applied to construct and analyse multimodal corpora of social media discourse. The case study focuses on the multimodal representation of #freerange on Twitter over a six-month period (Malory and Sha). The topic has gained prominence amid growing public concern in the UK about animal welfare and ethical consumption (RSPCA 2022), particularly in debates surrounding the ethics and transparency of free-range farming practices (Lekakis 2022).

The case study addresses three research questions: (1) How is #freerange usually represented - as food products or as animals? (2) How are associated social actors and free-range methods appraised? And (3) How does appraisal differ across tweeter identities? We will outline the data collection process and provide a detailed description of the corpus construction in the Methodology Section. Following this, we will introduce four functionalities for conducting multimodal corpus-assisted discourse studies using ATLAS.ti. This paper focuses on demonstrating the application of ATLAS.ti's functionalities for constructing and analysing multimodal corpora, highlighting their methodological value. For practical details of the software, such as operational steps and interface navigation, readers are encouraged to refer to the comprehensive manual available on the official ATLAS.ti website (ATLAS.ti GmbH 2024b).

3. Methodology

3.1. Data collection

The multimodal tweet dataset for (Malory and Sha) was collected using Twitter's "Advanced Search Function". The focus was on English-language tweets containing the hashtag #freerange within the period from 1st January, 2022, to 1st June, 2022. Data collection took place in July 2022, with January-June 2022 chosen as the period of focus due to an uptick in public debate about the ethics of free-range farming during the ongoing outbreak of avian flu affecting the UK (Race 2022). Increased discussion of free-range agriculture during this period, including on Twitter, provided valuable multimodal data for trialling the methodological protocol with ATLAS.ti outlined in this paper. Using a Google Chrome extension called "Twitter Media Downloader", we extracted both the tweet text and associated multimodal elements. During this process, the text, along with metadata such as usernames and the number of likes, was automatically saved in an Excel file. Other semiotic resources, including images and videos, were automatically downloaded in bulk and saved to a local folder. Following a manual review to filter out tweets irrelevant to the research focus, 1,003 tweets were retained.

As we were interested in comparisons across different tweeter identities, we further refined the dataset by retaining only tweets that could be categorised as commercial (primarily promoting products for financial gain), institutional (focused on animal rights advocacy), or personal accounts (lacking any discernible commercial or ethical affiliations). These classifications were inferred from the users' profile details and the content of their posts. However, it is important

to acknowledge the limitations of this categorisation method, especially in the case of personal accounts where profiles may be fictitious or misleading. As a result, additional exclusions of ambiguous cases were made, ultimately resulting in a final corpus of 899 tweets.

3.2. Constructing a multimodal corpus on ATLAS.ti

The initial step in constructing the corpus was uploading and setting up the target multimodal tweets. Importing the Twitter data as "Surveys" ensures the retention of self-selected metadata for analysis, such as Username, Tweet Content, and URLs. Once uploaded, each tweet is treated as an individual document within the ATLAS.ti project.

In the examined dataset, 94.22% of the tweets are multimodal, incorporating semiotic resources such as text, emojis, emoticons, memes, videos, external links and retweets. This high proportion highlights the importance of retaining multimodal features for social media studies.

Of these semiotic resources, text and emojis are directly displayed in the tweet cases. For other semiotic sources like memes, photos, and external links, the network function in ATLAS.ti allows for the linking of these resources within the same tweet. This enables the cohesive presentation of all semiotic elements on a single page, as demonstrated in Figure 1. To facilitate this, they need to be uploaded into the corpus using the "Add Documents" function and then linked to the corresponding tweet content (see Figure 1 for the multimodal pageview for a tweet case).

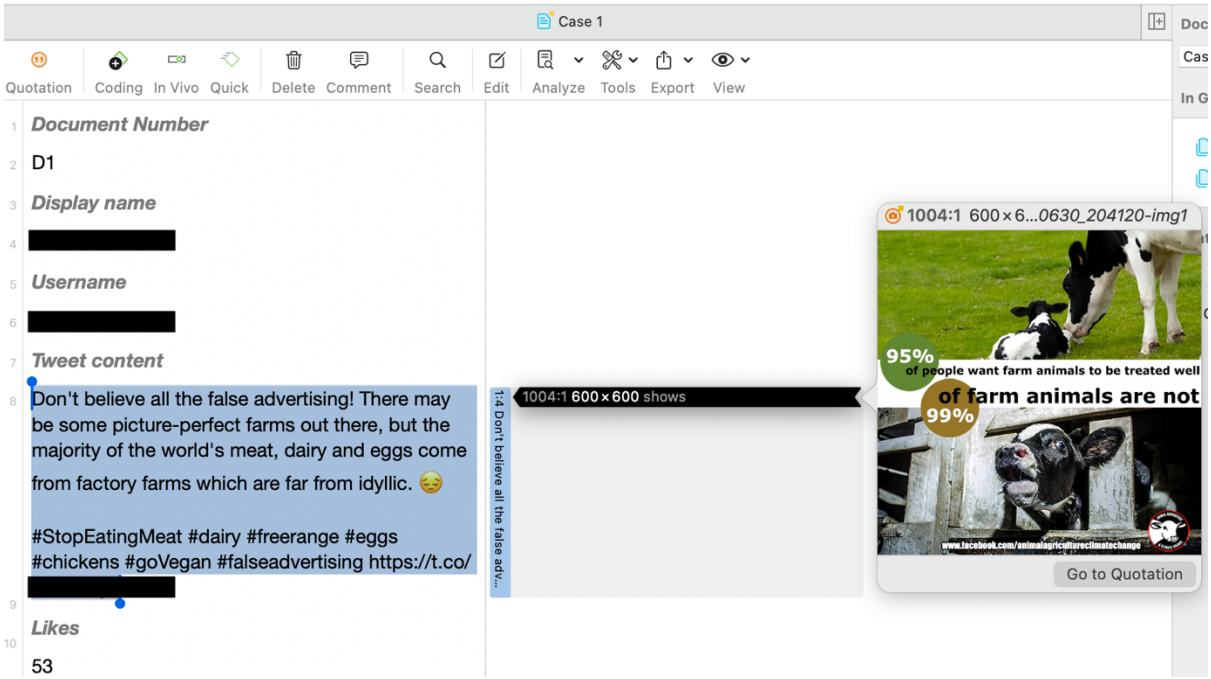

Figure 1. Pageview for a tweet case

Videos presented a unique challenge compared to other semiotic modes, primarily due to the labour-intensity associated with their analysis. For this study, we adopted a strategy of capturing screenshots from the initial meaning-making unit of each video. This is the frame most platform users see before interacting with the video, allowing us to analyse the videos as static images. Additionally, we carefully watched each video in full to ensure that no potential meaning conflicts, such as instances of irony, were missed.

3.3. Annotating multimodal corpora on ATLAS.ti

Since analysis and interpretation in ATLAS.ti rely on the "codes" (i.e., annotations) assigned to the data, the second step involves annotating the target corpus. Researchers have two options for this process. The first is the semi-automated "Intentional AI coding", which allows users to assign coding tasks based on their research questions. However, for MCADS, this method is currently only suitable for simple tasks such as detecting specific words or summarising general themes.

The second option is manual coding, which we employed in the present study. Manual coding provides researchers with greater flexibility to develop a coding scheme that aligns more with their specific research questions, rather than being restricted by the limitations of the software's analytical capabilities. The annotation process was conducted by the first author, with the second author reviewing the codes to ensure plausibility.

To address the research question "How does appraisal differ across tweeter identities?", we categorised user identities based on the criteria outlined in Section 3.1, classifying them into commercial, institutional, and personal accounts.

For the research question "How are associated social actors and free-range methods appraised?", we coded the tweets according to the appraisal strategies employed, with a particular focus on the Attitude dimension of Martin and White's (2003) Appraisal Framework (see Figure 2).

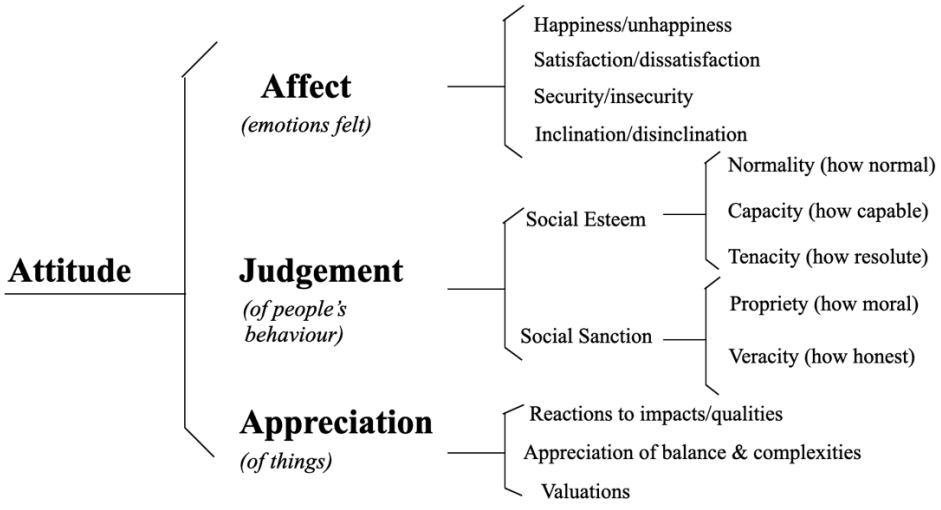

Figure 2. An overview of the Attitude dimension in Appraisal Framework (Martin and White 2003)

The Attitude coding involves not only the appraisal categories following a three-fold structure, such as [Appreciation-quality-positive] for Example 1, but also identifying the ideation (Inwood and Zappavigna 2024), or subjects of these attitudes (e.g., [free-range food] for Example 1). This allowed us to also address the first research question, which is to determine whether #freerange is usually represented as food products or as animals in the tweets.

Example 1:

Today's Lunch Feature: #freerange buttered chicken served with basmati rice and chapati!
Online order at [anonymised]

Additionally, we focused on the cross-modal synergy between appraisal strategies and social actor representations. Using a data-driven approach, we coded the visual representations of social actors into five categories: [free-range-few], [free-range-flock], [human commercial], [human consumer], and [human raiser]. This addresses the visual mode's prominent influence on audience perception, as evidenced by eye-tracking experiments (Garcia and Stark, 1991; Holmberg, 2004; Leckner, 2007). Specifically, such studies demonstrate that visual elements not only capture initial attention but also sustain the viewer's gaze for longer periods.

4. Analytical methods for MCADS on ATLAS.ti

Gillings et al. (2023) classify (monomodal linguistic) Corpus-assisted Discourse Studies (CADS) into two prototypical approaches, depending on whether Corpus Linguistics (CL) or Discourse Studies (DS) take the lead. In the first approach, CL drives the analysis by producing results that DS then interprets through detailed textual analysis, contextualising them within historical, sociocultural, and political frameworks. In the second approach, CL assesses the typicality of what DS first identifies in a small set of texts across larger corpora.

Our analysis of multimodal social media discourse follows the first approach, utilising ATLAS.ti's functionalities, such as (1) frequency analysis, (2) code co-occurrence analysis, and (3) the query tool, to initially identify statistically (non-)salient patterns. Researchers then analyse the associated (4) multimodal concordances and relate them to broader contexts for interpretation.

While the detailed discourse analysis is not as prominent here, this brevity is intentional, as our primary focus of this paper is methodological, i.e., demonstrating the application of MCADS using ATLAS.ti through a case study. The discursive aspect is explored more thoroughly in Malory and Sha. Here, however, our aim is to demonstrate the methodology itself.

4.1. Frequency analysis

In terms of frequency analysis, ATLAS.ti generates automatic statistics following the coding process, which can be accessed via the "Quotation Manager".

To illustrate its utility, we summarise the results on the visual representations of #freerange social actors in Figure 3. A comparison of tweeter identities reveals that #freerange is predominantly depicted as food across all three user groups. Furthermore, Figure 3 shows that

commercial accounts are more inclined to use visual representations of free-range food in their posts, compared to institutional and individual accounts.

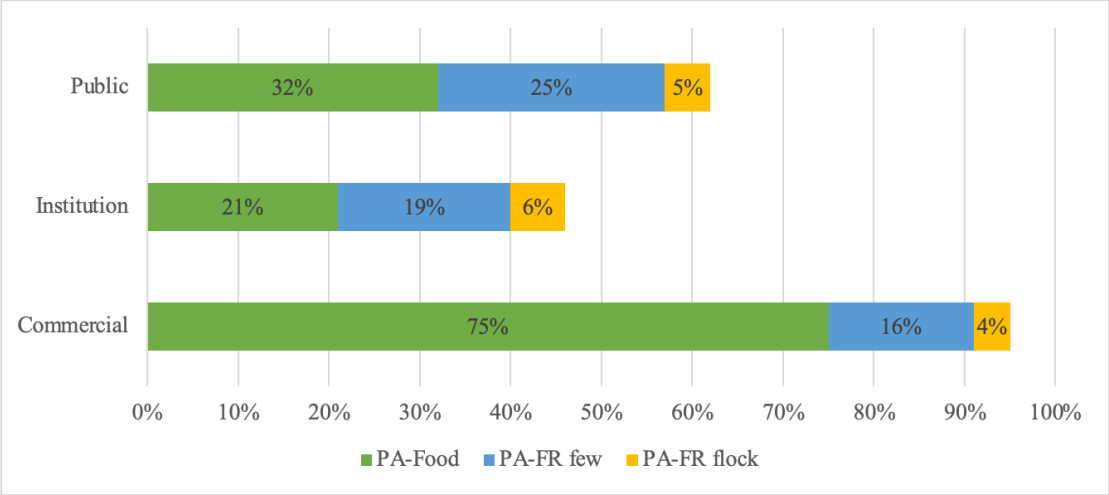

Figure 3. Visual representations of #freerange (by user identity)

4.2. Code co-occurrence analysis and associated multimodal concordances

4.2.1. Code co-occurrences & extended concordances

Code co-occurrence analysis enables the examination of co-occurring patterns of researcher-designed codes, such as themes, social actor representation strategies, and appraisal strategies in the present study, both within and across modes.

In addition to raw numbers of co-occurrences, ATLAS.ti also provides the co-occurrence coefficient, which is represented by the formula c = n12 / (n1 + n2 - n12), where n12 is the number of times two codes co-occur, and n1 and n2 are the occurrences of each individual code. The coefficient (c) indicates varying degrees of co-occurrence between the two codes. A value closer to 1 suggests a stronger association, with the codes frequently appearing together, while a value closer to 0 indicates a weaker relationship.

One illustration of code co-occurrence analysis can be found in Table 1, which shows the strengths of co-occurring patterns between visual representations of #freerange social actors and the attitudes expressed in the corresponding tweets.

	Free-range food/product	Free-range animals (few)	Free-range animals (flock)
Affect-happiness-positive	0.19	0.17	0.06
Affect-happiness-negative	0	0.03	0.06
Affect-inclination-positive	0.11	0.14	0.04
Affect-inclination-negative	0	0	0
Affect-satisfaction-positive	0.1	0.02	0
Affect-satisfaction-negative	0	0	0.04
Appreciation-balance & complexity-positive	0.16	0.42	0.07
Appreciation-balance & complexity-negative	0	0.02	0.3
Appreciation-impact-positive	0.31	0.04	0.01
Appreciation-impact-negative	0	0.02	0.03
Appreciation-quality-positive	0.72	0.06	0.01
Appreciation-quality-negative	0.02	0.01	0.01
Appreciation-valuation-positive	0.08	0.05	0.03
Appreciation-valuation-negative	0.01	0.04	0.15

Table 1. Coefficients between visual representations of #freerange social actors and appraised attitudes

The disparities among the appraisal of three #freerange social actor categories are evident in the results shown in Table 1. Firstly, tweets featuring visual representations of free-range food and/or products predominantly align with positive attitudes. The strongest association is observed in the co-occurrence coefficient of 0.72 between visual representations of free-range food or product and the appraisal of positive quality ([Appreciation-quality-positive]).

The differences become more intricate when considering the visual representations of free-range animals. Here, a critical distinction emerges between tweets depicting a few individual animals ([Free-range animals (few)] in Table 1) versus those showcasing a flock ([Free-range animals (flock)]). Tweets representing a small number of animals generally convey positive sentiments, as indicated by a co-occurrence coefficient of 0.42 with the positive appraisal of balance and complexity (i.e., [Appreciation-balance&complexity-positive]). In contrast, tweets portraying a flock of animals tend to express more negative appraisal, demonstrated by a notable co-occurrence coefficient of 0.30 with the [Appreciation-balance&complexity-negative] code.

Nevertheless, an inherent limitation of the coefficient and similar metrics is their susceptibility to distortion from disparate code frequencies. To guide researchers in identifying significant relationships that might otherwise be overlooked due to imbalanced application frequencies of codes, ATLAS.ti incorporates a visual cue: a yellow dot in the top right corner of the table cell. This marker appears when one code in a pair has been applied more than five times as frequently as its counterpart. This indicator prompts a closer examination of the co-occurrences within that cell (see ATLAS.ti GbmH. 2024b for details).

Once (non-)salient co-occurrence patterns have been identified, researchers can click on the corresponding co-occurrence columns, where ATLAS.ti displays the relevant "quotations" (i.e., concordances) of these co-occurrences on the right-hand side (see Figure 4). This structured pageview largely resembles the concordance functionality in monomodal linguistic corpus software. It also enables researchers to closely examine the multimodal contexts of target patterns, as well as the associated contextual factors of each concordance, such as engagement metrics and user account details (see Figure 1). The short-form nature of tweets, due to the character limit, makes them suited for multimodal concordance analysis, with each tweet functioning as its own concordance.

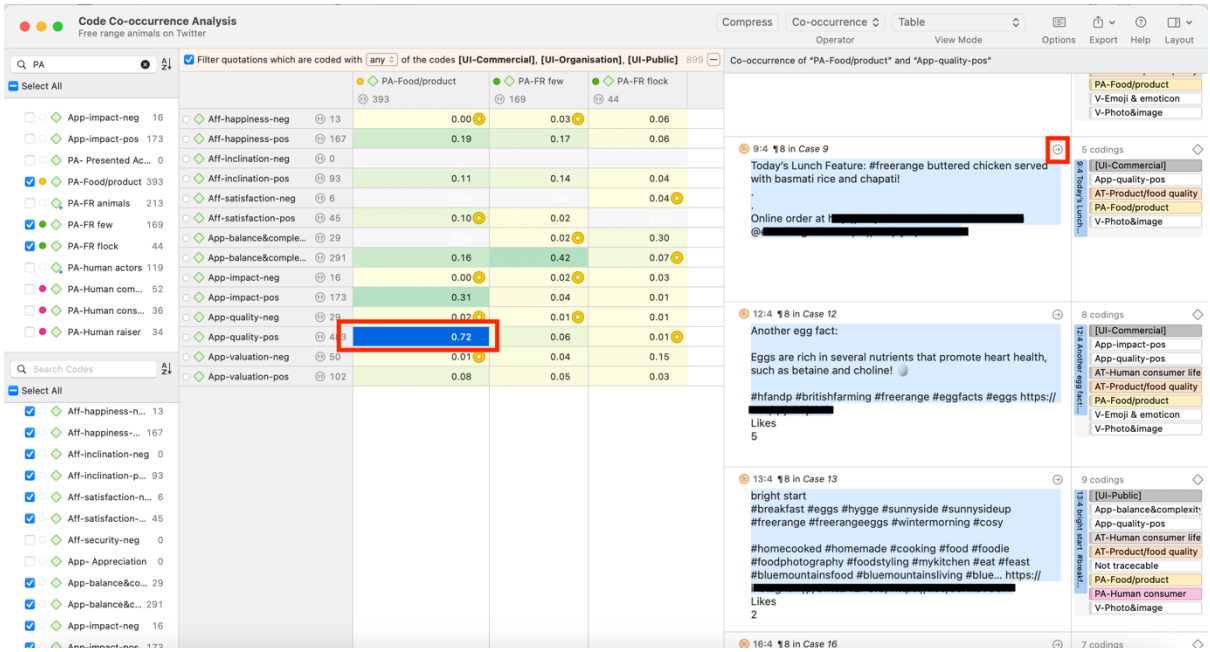

Figure 4. Code co-occurrence analysis and corresponding concordances

4.2.2. Visualisation

One of the strengths of ATLAS.ti in MCADS is its visualisation capabilities. One such visualisation tool is the Force-Directed Graph, which enables graphic representation of inter-code relationships. This method employs three elements to convey information: (1) size of node, suggesting the frequency of a code within the corpus; (2) length and (3) thickness of the connecting lines, where shorter, thicker lines suggest a higher coefficient and therefore stronger co-occurrence between codes. The corresponding multimodal concordances for these co-occurrences are also traceable, with a pageview similar to that shown in Figure 4.

This visualisation reveals the web of co-occurring relationships, uncovering patterns that might otherwise remain hidden. We now employ Figure 5, which shows the co-occurrence patterns across the appraised subjects in the corpus, to demonstrate the utility of Force-Directed Graphs.

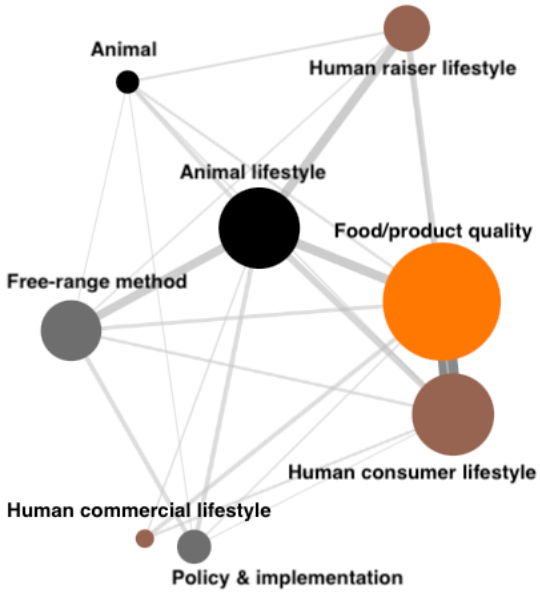

Figure 5. Co-occurring patterns across appraised subjects

The prominence of the "Food/product quality" node in Figure 5 highlights the evaluations of free-range food and products being the most frequent within the corpus. In contrast, the smallest "Human commercial lifestyle" node suggests its infrequent appearance in the evaluations. As previously outlined, the length and thickness of the lines in the figure illustrates the strength of co-occurrences between nodes. The shortest and thickest line, linking "Food/product quality" and "Human consumer lifestyle", suggests that these subjects are frequently evaluated together in the tweets. There are also notable connections between "Animal lifestyle" and "Food/product quality", as well as between "Animal lifestyle" and "Human consumer lifestyle" in Figure 5. However, some nodes remain disconnected, indicating a lack of co-occurrence among them within the dataset. For instance, the absence of a line between "Policy & implementation" and "Human raiser lifestyle" suggests that the associated evaluations never co-occur in the analysed corpus.

4.3. The Query Tool

The third functionality is the Query Tool, which can be used to filter concordances based on specific code (non-)co-occurrences. This tool is especially useful for further investigating particular (multimodal) discursive patterns within the corpus, or when unexpected patterns emerge that extend beyond the initial research questions, as is common in corpus studies (Gillings et al. 2022). Similar to the code co-occurrence analysis viewpage (see Figure 4), it also directs researchers to the corresponding concordances in detail and enables in-depth analyses of the patterns.

For instance, the code co-occurrence analysis summarised in Figure 5 reveals frequent pairwise co-occurring patterns across three subjects: animal lifestyle, food/product quality, and human consumer lifestyle. This prompted us to further investigate cases where these appraised subjects co-occur, especially in tweets from commercial accounts, and to explore whether causative links or sequential occurrences are represented.

We set the Query tool and qualitatively analysed the 15 cases that met the criteria, and found that 66.67% suggested a causal relationship; wherein a balanced and harmonious animal lifestyle is portrayed as directly contributing to high-quality food, which, in turn, is implied to support an environment- and health-conscious consumer lifestyle. This is illustrated in Example 2, which opens by emphasising the well-being of the cows, personifying them as "girls" and ascribing human emotions, "Happy Cows", in the text. The portrayal of the animals' lifestyle, via descriptions like "afternoon milking" and "a little massage", is positive, whilst the accompanying non-linguistic semiotic mode, a video which shows cows walking on green grass under a blue sky, further implies ease and leisure. The focus then shifts to milk quality, with the hashtag #farmfreshmilk implying a direct link between the cows' welfare and production of milk that is superior. Finally, the hashtag #lowfoodmiles ties this narrative to sustainable practices, highlighting reduced environmental impact and connecting animal welfare to the environmentally conscious and health-oriented lives of consumers.

Example 2

Happy Cows exited for their afternoon milking, and having a little massage on the way in #happycows #freerange #farmfreshmilk #sustainablefood #lowfoodmiles

5. Conclusions

In this paper, we have outlined a novel approach to constructing and analysing multimodal corpora of short-form social media discourses using ATLAS.ti. For multimodal corpus construction, ATLAS.ti facilitates the integration of multimodal elements from social media platforms, enabling the demonstration of multimodal concordances within single pageviews, along with broader contextual information, such as engagement metrics and user details.

For MCADS, we have demonstrated how three functionalities, i.e., frequency analysis, code co-occurrence analysis (including visualisation), and the query tool, can be utilised to identify (non-)salient patterns of mono- and multimodal code (non-)co-occurrences. These tools provide a macroscopic view of the corpus while addressing the critical challenge of understanding the strength and the underlying reasons for patterns of cross-modal co-occurrences. As illustrated, the Force-Directed Graph and other visualisation tools enhance this analytical process by providing intuitive representations of complex patterns.

The potential of ATLAS.ti for MCADS is further demonstrated by its ability to provide easy access to multimodal concordances through the viewpages associated with each analysis function. This enables a detailed examination of the co-occurrence patterns within each concordance. Additionally, the contextual information available on the viewpage is valuable for conducting in-depth analyses of the concordances.

Working with ATLAS.ti for MCADS does present certain challenges, primarily related to the investment of time required for corpus construction and annotation. Factors such as the coder's familiarity with the software and the coding scheme for their research can influence the pace of work considerably. In our pilot study of 50 tweets, the initial stage of corpus construction and coding averaged approximately 5 minutes per tweet, as we were still familiarising ourselves with the process. However, this pace accelerated significantly during the main study due to increased familiarity with the tools, and a structured scheme can lead to greater efficiency in subsequent phases.

To address these challenges, researchers may find it beneficial to start with a pilot study using a smaller but representative data sample, to become acquainted with the software and refine the details of the coding scheme. Further streamlining of the process can be facilitated by pre-

organising media files to align with their corresponding text files (e.g., through consistent naming conventions) and by utilising ATLAS.ti's bulk document upload feature. Additionally, the Quotation Manager page provides a structured overview and simplifies linking and tracking processes. For corpus annotation, implementing a tiered coding scheme, such as the three-fold structure used in this study (see Section 3.3) can greatly enhance both organisation and efficiency.

As illustrated in the case study presented in this article, the triangulation approach of integrating ATLAS.ti's functionalities, such as the automatic quantification of multimodal (non-)co-occurrences, with detailed, qualitative manual analysis of the associated concordances, enables the identification of patterns that might otherwise have gone unnoticed. In this case study, examples include tweeters' preference for visually representing #freerange with food rather than animals, as well as the causal relationships underlying commercial accounts' positive evaluations of #freerange, specifically progressing from animal lifestyle to food quality and, ultimately, to human consumer lifestyle. In this way, by utilising ATLAS.ti for MCADS, this methodology provides a meaningful contribution to advancing research in the field.

References

Allwood, Jens. 2008. *Multimodal corpora*.

Andersson, Marta. 2024. Multimodal expression of impoliteness in YouTube reaction videos to transgender activism. *Discourse, Context & Media*, *58*, 100760.

Antoniadou, Victoria. 2017. Collecting, Organizing and Analyzing Multimodal Data Sets: The Contributions of CAQDAS. *Research-publishing. net*.

ATLAS.ti GmbH. 2024a. ATLAS.ti Feature Comparison. https://atlasti.com/feature-comparison

ATLAS.ti GbmH. 2024b. ATLAS.ti 22 User Manual. https://doc.atlasti.com/ManualMac.v22/Intro/IntroductionPreface.html

ATLAS.ti Scientific Software Development GmbH. 2024. ATLAS.ti Mac (version 24.2.0) [Qualitative data analysis software]. https://atlasti.com

Baker, Paul, Costas Gabrielatos, Majid Khosravinik, Michał Krzyżanowski, Tony McEnery, and Ruth Wodak. 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & society*, *19*(3), 273-306.

Baker, Paul, & Erez Levon. 2015. Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication*, *9*(2), 221-236.

Baker, Paul, & Jesse Egbert. 2016. *Triangulating methodological approaches in corpus linguistic research*. London: Routledge.

Baker, Paul & Luke Collins. 2023. Creating and analysing a multimodal corpus of news texts with Google Cloud Vision's automatic image tagger. *Applied Corpus Linguistics*, *3*(1), 100043.

Baker, Paul, & Tony McEnery. 2005. A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of language and politics*, *4*(2), 197-226.

Baker, Paul, & Tony McEnery. 2015. Who benefits when discourse gets democratised? Analysing a Twitter corpus around the British Benefits Street debate. In *Corpora and discourse studies: Integrating discourse and corpora*, 244-265.

Bednarek, Monika, & Helen Caple. 2017. Introducing a new topology for (multimodal) discourse analysis. *Transforming Contexts*, 19.

Bouvier, Gwen, & David Machin. 2020. Critical discourse analysis and the challenges and opportunities of social media. *Critical discourse studies and/in communication*, 39-53.

Caple, Helen. 2018. Analysing the multimodal text. *Corpus Approaches to Discourse: A Critical Review*, 85-109.

Chałupnik, Małgorzata, & Gavin Brookes. 2022. Discursive acts of resistance: a multimodal critical discourse analysis of All-Poland Women's Strike's social media. *Gender & Language*, *16*(3).

Christiansen, Alex, William Dance, & Alexander Wild. 2020. Constructing corpora from images and text. *Corpus approaches to social media*, 149-174.

Choe, Yunseon, Jiyoon Lee, & Gyehee Lee. 2022. Exploring values via the innovative application of social media with parks amid COVID-19: A qualitative content analysis of text and images using ATLAS. Ti. *Sustainability*, *14*(20), 13026.

Clarke, Isobelle. 2022. A Multi-dimensional analysis of English tweets. *Language and Literature*, *31*(2), 124-149.

Collins, Luke. 2020. Working with images and emoji in the Dukki Facebook Corpus. *Corpus Approaches to Social Media. Amsterdam: John Benjamins*, 175-96.

Djonov, Emilia, & Sumin Zhao. 2013. From multimodal to critical multimodal studies through popular discourse. In *Critical multimodal studies of popular discourse*, 13-26.

Friese, Susanne, Jacks Soratto, and Denise Pires. 2018. Carrying out a computer-aided thematic content analysis with ATLAS. ti.

Garcia, Mario, & Pegie Stark. 1991. *Eyes on the news*. Poynter Institute.

Gillings, Mathew, Gerlinde Mautner, & Paul Baker. 2023. *Corpus-assisted discourse studies*. Cambridge University Press.

Han, Joshua, & Michele Zappavigna. 2024. Multimodal rhythm in TikTok videos: Exploring a recontextualization of the Gillard 'misogyny speech'. *Multimodality & Society*, *4*(1), 58-79.

Hansson, Sten, & Ruth Page. 2023. Legitimation in government social media communication: The case of the Brexit department. *Critical Discourse Studies*, *20*(4), 361-378.

Holmberg, Nils. 2004. Eye movement patterns and newspaper design factors: An experimental approach.

Inwood, Olivia, & Michele Zappavigna. 2024. The legitimation of screenshots as visual evidence in social media: YouTube videos spreading misinformation and disinformation. *Visual Communication*, 14703572241255664.

Jovanovic, Danica, & Theo Van Leeuwen. 2018. Multimodal dialogue on social media. *Social Semiotics*, *28*(5), 683-699.

Leckner, Sara. 2007. An eye-tracking study of the influence of media technology on behaviour patterns in newspaper reading.

Lekakis, Eleftheria. 2022. *Consumer activism: Promotional culture and resistance*. Sage.

Malory, Beth, & Sha Yuze. Forthcoming. Ethical consumerism or symbolic capitalism? Public multimodal evaluations of #freerange on Twitter.

Martin, James, & Peter White. 2003. *The language of evaluation* (Vol. 2). Springer.

Ng, Chi Wei. 2018. Critical multimodal discourse analyses of news discourse on Facebook and YouTube. *Journal of Asia TEFL*, *15*(4), 1174.

O'Halloran, Kieran. 2011. Investigating argumentation in reading groups: Combining manual qualitative coding and automated corpus analysis tools. *Applied linguistics*, *32*(2), 172-196.

Page, Ruth. 2022. Analyzing multimodal interactions in social media contexts. *Research Methods for Digital Discourse Analysis*, 159.

Paulus, Trena, & Jessica Lester. 2014. *Using ATLAS. ti for a discourse analysis study of blogging in an educational context*.

Potgieter, Andrea. 2014. Learning through teaching: ATLAS. ti and Social Media. https://doi.org/10.14279/depositonce-4848.

Race, Michael. 2022. Free-range eggs no longer available in UK due to bird flu. *BBC News*. https://www.bbc.co.uk/news/business-60820595

RSPCA. 2022. *Animal Kindness as part of our values*. https://www.rspca.org.uk/whatwedo/latest/kindnessindex/report

Widdowson, Henry. 2000. On the limitations of linguistics applied. *Applied linguistics*, *21*(1), 3-25.

Woods, Megan, Trena Paulus, David Atkins, & Rob Macklin. 2016. Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS. ti and NVivo, 1994–2013. *Social science computer review*, *34*(5), 597-617.