



# Scraping the Monumental: Stepan Bandera through the Lens of Quantitative Memory Studies

---

ROLF FREDHEIM, GERNOT HOWANITZ, MYKOLA MAKHORTYKH

*University of Cambridge, University of Passau, University of Amsterdam*

**Abstract:** In this article we use the example of Stepan Bandera to demonstrate the effectiveness of web-scraping methods as a tool to explore how people interact with memory content online. Using data from *Wikipedia*, *Twitter* and *YouTube*, we analyse the traces left by users interested in Stepan Bandera and assess how these differ between Ukraine, Russia and Poland. Applying data mining and content analysis to data traditionally analysed from a purely qualitative perspective, we show how Polish content about Bandera follows completely different patterns in comparison to Ukrainian and Russian debates. Our tools, made available online, also include attempts at analysing video content.

**Keywords:** Wikipedia, Twitter, YouTube, memory studies, data mining, content analysis, digital humanities, Eastern Europe

---

Who is Stepan Bandera? The answer – as often happens with controversial historical figures – depends on whom you ask. Bandera, a leader of the Ukrainian nationalist movement during the Second World War, was already controversial during his lifetime. In the Soviet Union he was branded as a Nazi collaborator; Ukrainian émigrés praised him as a fearless independence fighter.<sup>1</sup> The existing scholarship features a variety of interpretations: from Nazi collaborator, terrorist and instigator of ethnic cleansings to national hero, fearless resistance leader and martyr<sup>2</sup>. Consequently, it is hard to disagree with David Marples, who argued that the name of ‘Bandera’ provokes an emotional reaction simply by its utterance (Marples 2006: 555). In this paper we show that Marples’ assertion is supported not only by

---

<sup>1</sup> We collected the data for this study mostly in March and April 2013. Therefore, we could not take the recent developments in the relationship between Ukraine and Russia – Euromaidan and the annexation of the Crimea peninsula – into account. It is quite likely that online activity related to Stepan Bandera has shifted since last year.

<sup>2</sup> To mention some influential works: Hordasevych (2001); Prus (2004); Marples (2009); Balyns’kyi, Amar and Hrytsak (2010).

offline controversies – including international demarches, desecration of monuments and heated public debates – but also by online debates across various social media platforms.

This article provides a toolkit that demonstrates how data mining and content analysis may be applied to content traditionally analysed from a purely qualitative perspective. As a field, Memory Studies has been relatively slow to embrace quantitative methods. Researchers have employed quantitative methods to trace the number of existing texts about a subject at various junctures but while this tool can identify levels of interest, these same researchers struggle to unpack the meaning of these interest levels without laboriously studying individual texts. While this article says little new about Stepan Bandera, digital methods, or memory studies individually, it does demonstrate that a lens combining the three can be focused and adapted to answer questions about how ideas are discussed online.

To conduct this study, we employed web-scraping techniques to harvest user-generated online data about Stepan Bandera. By scraping not just content, but also user patterns, time series analyses can be applied to determine when specific groups interact with a subject. Analysing *YouTube* data both sheds lights on the content of videos and indicates how and when users post and watch them. Finally, *YouTube* comments created by users who communicate in a particular language can be treated as a single text, which allows diverging interests to be mapped in broad brush strokes.

In 2009, as Ukrainian nationalists passed through Poland during a Bandera march to Munich, the location of Bandera's grave, the Polish *Wikipedia* page about Stepan Bandera exploded into life, with the average daily number of visits growing from roughly 50 to 800. Since then, Polish interest in Bandera, as expressed through *Wikipedia* page views, has consistently been at least twice the level prior to 2009, even when controlling for increased internet usage<sup>3</sup>. The period of increased interest coincided not only with concrete action undertaken by Bandera's Ukrainian supporters, but also with a period of increased Polish awareness of the Volhynia massacre. On 27 October 2012, the day before the Ukrainian parliamentary elections, Polish social media sites again came to life, as a virulently aggressive rap video, rated as 18+ on *YouTube*, spelled out the connections between Volhynia, Bandera, atrocities committed against Polish citizens in the east and the present Ukrainian political climate. We detected no analogous upsurge in Russian user-generated content on *YouTube* in 2009, and the number of users accessing the Russian language *Wikipedia* page was stable and moderate also in 2012. The topics of Bandera and Volhynia again flared up during the 70th anniversary of the massacre in July 2013, when the Polish Sejm declared the Volhynia massacres to be 'ethnic cleansing bearing the hallmarks of genocide'. Unsurprisingly, *Wikipedia* interest in Poland grew again. Ukrainian and Russian interest, on the other hand, dropped at these junctures. This brief introductory example hints at how the Ukrainian-Russian debate about Bandera is disconnected from the narrative about Volhynia in the Polish-language section of the internet. It further suggests how *Wikipedia* and *YouTube* are used differently and

---

<sup>3</sup> Throughout we have used adjusted values relative to the percentage of the population with internet access. Average growth rates were estimated based on World Bank data: <http://data.worldbank.org/indicator/IT.NET.USER.P2/countries?display=graph>, (accessed 19 November 2013). Adjustment is important, because otherwise it looks like all trends are increasing, when in reality the trajectory is unsurprising due to a growing number of internet users.

the complex ways offline actions may be accompanied by an online reaction. These reactions are sporadic, but when they occur they leave a substantial trace.

Alexander Etkind describes how memory mediated online is contextualised and made meaningful through memory models, where simple analogies are drawn between the past and the present (Etkind 2013). In online interactions, the interplay between memory events and memory models is mediated by internet prosumers who through their consumption of memory help to spread it: *YouTube* videos with more views and likes, especially within a short time frame, are more likely to go viral, thus breaking out of a restricted online community, to reach a larger part of the population. Prosumers, or memory users, select from available memory models when promoting particular news stories on *Twitter*, creating ‘demotivators’<sup>4</sup> in Photoshop, or even releasing video content on *YouTube*. In all these categories a distinction may be made between content re-appropriation and creation: in the former, offline content is selected, possibly edited or contextualised and spread online. Examples include television documentaries, newspaper articles, etc. This content is edited by users who decide how to frame it: within a tweet, with a particular interpretation, or with a given title on *YouTube*. In user-created content, evocative visual symbols are accompanied by a narrative linking the past to the present.

Usage patterns may be discerned by aggregating information about when videos were accessed, when *Wikipedia* edits took place and when users tweeted about Bandera. Patterns of activity are generally predictable, as memories are often invoked in cycles, for instance every year at anniversaries, or coinciding with the political calendar (Fredheim 2013a). Disproportionate increases at these critical junctures point to the viral spread of information, caused by either offline or online commemorative activities. Similarly, large jumps in online activity occurring between these junctures point to possible ruptures in the memory narrative caused by memory events. Memory events are ‘a re-discovery of the past that creates a rupture with its accepted cultural meaning [...] [and therefore] secondary to the historical events they interpret, usually taking place many years or decades later’ (Etkind 2010: 4).

Online social network platforms constrain the type of activity that is possible. Different platforms impose various constraints, but considerable similarities exist across networks. These may be explained by the special dynamics of online genres.<sup>5</sup> On *Facebook*, for instance, users operate under their offline identity and, typically, interact with people they have previously met in person. Both factors are to a lesser degree true for *Twitter*: many users link their *Twitter* profile to their offline identity, but this is by no means compulsory and it is also easier to ‘follow’ strangers. While users (often mistakenly) assume *Facebook* is private, tweets are explicitly public. Tweeted content is linked to other users through retweets<sup>6</sup>, direct messages and hashtags. The use of these social hyperlinks creates loose communities of *Twitter* users. On *YouTube*, in contrast, no such communities exist, as content is equally available to all users, assuming that language does not form a barrier. Online interactions are usually brief; the blog-post essay is the exception rather than the rule, and on *Twitter*, *YouTube* and *Wikipedia* discussions, complex questions of historical truth are frequently re-

---

<sup>4</sup> This image genre, especially popular in Eastern Europe, juxtaposes an image and a cynical or ironic statement.

<sup>5</sup> A comprehensive discussion of online genres by Dirk Uffelmann is part of this issue of *Digital Icons*.

<sup>6</sup> Retweeting allows *Twitter* users to forward and repost tweets.

duced to simple repeatable patterns, drawing heavily on memory models, analogies and symbolism, as epitomised by iconic images of Stepan Bandera, national emblems and flags etc., which are remediated in user-generated content. By making these patterns the subject of our study, we aim to expose how people refer to Stepan Bandera online. Just as this symbolism may be discerned by quantifying textual patterns, it may also be accessed by automated genre detection, e.g. by distinguishing user-created slide-shows and filmed content. By using keyword analysis, the disproportionate presence of these symbols within different online genres and language communities may be identified and contrasted with those of other communities, as well as with the usually more politically correct text agreed upon by *Wikipedia* editors.

The processes involved in collecting, aggregating and quantifying time series and textual and visual data are complex and are discussed elsewhere. Throughout this article we attempted to avoid highlighting ‘medium specific features’ of the data and ‘data centrism’ by relegating technical details and the minutiae of evidence to footnotes and external links (Marres 2013). As one main goal of this article is to demonstrate which data are available across different online platforms and how this data can be further processed, we made both data and the scripts used to create our evidence available for download, meaning that anyone can reproduce our findings,<sup>7</sup> or adapt our methods to other topics.

This article is divided into six sections. First, we formulate our hypotheses. In the following three sections we describe the scraping process in detail for each of the three online platforms: *Wikipedia*, *Twitter* and *YouTube*. In the fifth section we discuss our findings. Finally, in the concluding section, we check the validity of our hypotheses.

## Hypotheses

- We hypothesised that the language barrier between Western Slavic Polish on the one hand, and Eastern Slavic Russian and Ukrainian on the other, is substantial. We therefore expected to see Poland as a special case in every regard.
- We hypothesised that language trumps location. In other words, we hypothesised that the possibility of a conflict concerning Bandera’s personality is higher between speakers of Russian and Ukrainian, rather than between people living in Russia and people living in Ukraine.
- We ventured that Poles and Russians would create the most hostile or negative content about Bandera.
- We expected predictable patterns of interest that coincide with annual anniversaries and Ukrainian political elections. We expected that these patterns would be similar across various online media and genres.

---

<sup>7</sup> The scripts and data used may be downloaded from <https://github.com/ghowa/bandera-scripts> (accessed 1 December 2014).

## Bandera on Wikipedia

*Wikipedia* is the most popular online encyclopaedia, with its vast number of articles composed by internet users. In the context of our analysis, the process of writing and reading *Wikipedia* articles is of as much interest as their content. This platform allowed us to compare how the different language versions of the Stepan Bandera article relate to each other. Our analysis was based on the Russian, Ukrainian, Polish and English versions<sup>8</sup> and employed three main approaches: a close reading and a distant reading<sup>9</sup> of the articles as well as some thoughts about the articles' editors. We relied on the statistical data which is available on the *Wikipedia* website. In Figure 1, we plot the number of page views over time. Relative to the size of the Internet population, the Ukrainian page has seen the largest number of visits. Generally, Polish and Russian interest is remarkably consistent, but two differences are notable: both in 2009/10 and in 2012/13 the Polish and Russian trajectories diverged strongly: Russian interest was focused squarely on the 'Bandera marches' and the Ukrainian legislation, according to which Bandera was proclaimed a hero of Ukraine. In 2010, Poles also read *Wikipedia* at the time of the Ukrainian elections. However, it was mainly users of the Polish *Wikipedia* who read about Bandera on 8 August 2009 when Ukrainians cycling Bandera's route to Munich were stopped at the Polish border and denied access to Poland,<sup>10</sup> or during the 70-year commemorations of the Volhynia massacre in July 2013. First and foremost, Ukrainian interest has been highly predictable: every year in January at the time of Bandera's birthday, there was an upsurge in interest. This pattern, however, is not matched by Polish or Russian internet users (see Figure 1).

### *Close reading Wikipedia*

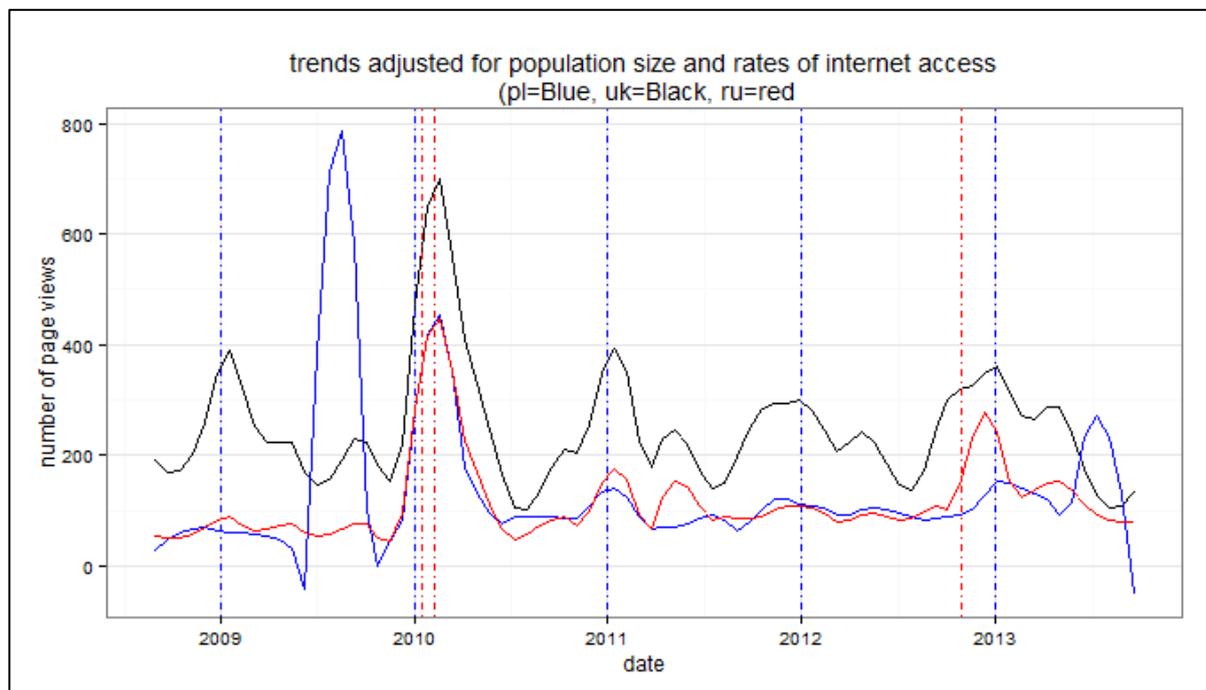
Applying close reading to large corpora of multilingual data is difficult, due to both the quantity and quality of data. One possible solution is to read only selected parts of articles, which are common to all language versions. This strategy is applicable to *Wikipedia* articles, because the different language versions are built on a common default template, which includes a lead section, a body section and standard appendices. Lead sections are especially interesting for a close reading, because they are rather short and function as a brief summary for the whole article. They also include so-called 'Persondata' templates, a box holding the most relevant information such as name, date of birth, citizenship, etc.

---

<sup>8</sup> [http://ru.wikipedia.org/wiki/Бандера,\\_Степан\\_Андреевич](http://ru.wikipedia.org/wiki/Бандера,_Степан_Андреевич), [http://uk.wikipedia.org/wiki/Бандера\\_Степан\\_Андрійович](http://uk.wikipedia.org/wiki/Бандера_Степан_Андрійович), [http://pl.wikipedia.org/wiki/Stepan\\_Bandera](http://pl.wikipedia.org/wiki/Stepan_Bandera) and [http://en.wikipedia.org/wiki/Stepan\\_Bandera](http://en.wikipedia.org/wiki/Stepan_Bandera) (all accessed 29 March 2013). English *Wikipedia* is listed here because English is the lingua franca of Internet communication and is therefore often employed by users of different nationalities, including Eastern Europeans. Consequently, the English version of the encyclopedia can be used as a neutral ground by Polish, Russian and Ukrainian users, where local discussions can be revived and brought to the attention of the Anglophone Internet audience.

<sup>9</sup> Franco Moretti coined the term 'distant reading' to propose a quantitative approach to literature 'where distance [...] is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes – or genres and systems' (Moretti 2000: 57, italics original).

<sup>10</sup> [http://wyborcza.pl/1,76842,6907784,Szlaban\\_na\\_rajd\\_Bandery.html](http://wyborcza.pl/1,76842,6907784,Szlaban_na_rajd_Bandery.html) (accessed 19 November 2013).

**Figure 1.** *Wikipedia* page views of the Russian, Ukrainian and Polish Bandera articles

Source: Rolf Fredheim

A division between English/Ukrainian articles about Bandera on the one hand, and Russian/Polish ones on the other, was immediately clear. While all four lead sections characterised Bandera as a political activist and representative of the Ukrainian nationalist movement, only the Ukrainian article described him as a ‘prominent’ figure. The Russian and Polish pieces, on the other hand, emphasised connections between Bandera’s views and fascist movements. This observation regarding division found further support in the description of Bandera’s historical role in the (attempted) proclamation of an independent Ukraine on 30 June 1941, which is stressed in the Ukrainian and English versions. The Polish and Russian articles, in contrast, emphasised Bandera’s involvement in a series of ‘terrorist acts’ (Russian article) and his involvement in the Volhynia massacre of Poles (Polish article).

The differences were also present in the Persondata template. All four articles used the same picture of Stepan Bandera and agreed on the basic details – date of birth, date of death, etc. Other aspects were less harmonious. For instance, the Ukrainian Persondata section deliberately mentioned that Bandera was born on the territory of contemporary Ukraine and was a Ukrainian by nationality. It also displayed the red-black flag of the Ukrainian Insurgent Army (Ukrains’ka Povstans’ka Armiia, UPA) and noted that Bandera was the first head of the Revolutionary Organisation of Ukrainian Nationalists. The English section provided a similar representation, stating the allegiances of Bandera to Ukraine and UPA through the use of miniature blue-yellow and red-black flags and describing his nationality as Ukrainian. The Russian Persondata section mentioned Bandera’s Ukrainian nationality and stressed that he was born on the territory of contemporary Ukraine, but added that he had Polish citizenship. The Persondata for the Polish article was the simplest of all four sections and did not

include any information other than the dates of birth and death. This observation suggests that the image of Bandera in the Polish *Wikipedia* is the least contested among all three versions of the encyclopaedia.

A third important source of information in this regard is reflected in the categories by which *Wikipedia* articles are classified. According to *Wikipedia*'s definition, '[...] [t]he central goal of the category system is to provide navigational links to all Wikipedia pages in a hierarchy of *categories* which readers [...] can browse and quickly find sets of pages on topics that are defined by those characteristics.' (Wikipedia 2013, italics original). Categories provide the semantic context of an article inside the particular language version; therefore, their comparison can provide a basic perspective on the semantic differences among different articles (Voss 2006). In order to facilitate an analysis we grouped categories based on their subject matter. These metacategories were: biographical data (place/date of birth and death), life activities (education/activities/hobbies), circumstances of death, commemoration (posthumous awards/landmarks) and related concepts (general notions/general associations). While all four articles shared the same metacategories,<sup>11</sup> there were a number of important semantic differences on the level of single metacategories. The 'Life Activities' metacategory was of particular interest in this case: all four articles categorised Stepan Bandera as a nationalist politician; however, the Polish article identified him as a collaborator with the Third Reich while the English article listed Stepan Bandera among Ukrainian anti-Communists, as does the Ukrainian article. It is interesting to note that the Polish text was keen to distance Bandera from anti-Communism – a stance that is predominantly positive in Polish political rhetoric. Together these differences point to the fact that instead of a presumably neutral point of view, *Wikipedia* offers a 'linguistic point of view' (Massa and Scrinzi 2013) which varies from one language version of the encyclopaedia to another.

### *Distant reading Wikipedia*

In order to compare the results of the close reading of the lead sections of articles with the main text from the body sections, we computed the word frequencies of the articles. The method used here is also used below for *YouTube* comments and *Twitter* data, and it is therefore described in brief only once. We translated Ukrainian, Russian and Polish text with Google Translator to facilitate the comparison between the different language versions and removed very common words using a list of stop words. The use of Google Translator is appropriate in this case, because, although imperfect, our results can be replicated. Certainly, there were a number of inaccuracies, e.g. incorrect interpretation of specific terms and names – for instance, the surname *Bandera* was sometimes translated from Polish as 'flag'. The results were clear, though: whereas both the Ukrainian and English texts included terms such as 'hero', 'honorary' and 'citizen', those terms were quite simply absent in the Russian and Polish versions. Instead, the Polish and Russian articles placed greater emphasis on words such as 'arrest', 'death', 'murder', 'concentration' and 'camp'. The fact that these negative terms were mostly absent from the English and Ukrainian texts might indicate that they con-

---

<sup>11</sup> The only exception to this rule is the Polish article which does not include any 'related concepts'.

veyed a more positive interpretation of Stepan Bandera. The English version paid less attention to Bandera's death, preferring to highlight how the historical figure is remembered. This was reflected through the frequency of terms such as 'monument', 'street' and 'museum'. In this way, Stepan Bandera is presented as a timeless symbol that is still relevant today. This stance resonates with the popular Bandera-related meme 'Heroi ne vmyraiut' ['Heroes never die'].

Use of the tool *Textviz* confirmed that many of the differences exhibited in the main articles were statistically significant.<sup>12</sup> Of the four corpora, the Ukrainian used the word 'hero' most frequently. In comparison with the other articles, it also indicated that the greatest emphasis on posthumous commemoration of Stepan Bandera is in contemporary Ukraine – particularly, the awarding of the title of Hero of Ukraine. This is evident in the use of such words as 'title', 'administration', 'president' and 'decision'. While the English discussion also paid considerable attention to the contemporary commemoration practices, more often it used words such as 'monument' or 'street', which were related to different sorts of practices. Furthermore, the Russian and Polish corpora differed from the Ukrainian view on Bandera's personality. The Polish article emphasised such words as 'OUN' (the English equivalent of 'UPA') and 'arrest', but also 'kill' and 'L'viv'.

Another approach towards our distant reading of *Wikipedia* articles was to scrape the electronic sources cited in the individual *Wikipedia* articles, i.e. all outgoing links. Those references represent an important part of a *Wikipedia* article, as they support the main arguments. Our findings indicated significant intersections between the Russian and Ukrainian versions, which had 17 sources in common. The Russian and Polish articles shared only one common source, and there were four common sources between the Ukrainian and Polish versions. Only two links were featured in all three articles, namely those linking to MediaWiki and to the Simon Wiesenthal Center. The Polish narrative about Bandera thus immediately stands in contrast to the Ukrainian and Russian ones, which share a considerable number of sources and, therefore, are more likely to represent at least two interpretations of Bandera's personality (namely, Ukrainian and Russian) in the same article.

### *Behind the scenes: about the editors*

*Wikipedia* articles are written collaboratively by Internet users, so we explored who makes the edits, and in particular, where they are located. As the *Wikipedia* privacy policy prohibits tracing the geographical location of registered editors, we were able to access this kind of data only for anonymous users. We applied a series of utilities made available by the University of Amsterdam's Digital Methods Initiative to access *Wikipedia* data. The *Wikipedia Edits Scraper* tool allowed us to gather data about registered and anonymous editors – namely the nicknames of the former and the IP addresses of the latter – and to compare these data for

<sup>12</sup> *TextViz*, a bit like a Word Cloud, plots the most frequently seen words. The strength of *TextViz* is that it plots not just frequency, but statistically significant differences between texts, and thus uses space in a meaningful way. See: <http://quantifyingmemory.blogspot.co.uk/2013/02/textviz.html> (accessed 19 November 2013).

three *Wikipedia* articles.<sup>13</sup> All in all, the Russian version had 471 different editors (271 anonymous), the Ukrainian 144 (33 anonymous) and the Polish 98 (58 anonymous).<sup>14</sup> The Russian and Ukrainian version shared 16 common editors, the Polish and Ukrainian 1, and the Russian and Polish 2. Again there was a clear pattern – very few people working on the Ukrainian or Russian site contributed to the Polish debate about Bandera.

After this initial comparison we used the *WikiTrip* tool in order to visualise the geographical location of an article's editors.<sup>15</sup> The results of this analysis indicated essential differences between the anonymous editing of the Polish article, on the one hand, and the Ukrainian and Russian articles on the other. Of all three articles, only the Polish text indicated a clear correlation between the language version of the *Wikipedia* article and the geographical location of editors. Of all edits of the Polish article, 88.37% were made from Poland. The Ukrainian article received the majority of its anonymous edits from Ukraine (46.59%), though there were several edits from the Russian Federation (3.07%), the United States (13.84%) and Poland (6.17%). The Russian *Wikipedia* article indicated the largest anonymous activity from Ukraine (54.57%) with the Russian Federation (37.5%) in second place and a few edits from the United States (1.82%), Kazakhstan (0.6%) and few other countries. The Russian article contained no anonymous edits from Poland.

The next step was to implement a sentiment analysis test on the collections of comments from these discussion pages. This approach was based on checking individually whether a word is present in a list of 4,000 adjectives classified as negative according to the word polarity dictionary used by Hu (2004). The fact that the texts were translated by machine means that errors are significant as typos and swear words remain unidentified. *A priori*, we expected the English language texts to appear the most negative, given that no translation is needed; conversely we expected the Ukrainian ones to appear the most balanced, given that Google translates Ukrainian by proxy of Russian. Across the board, nearly 1 in 20 words were identified as negative by the algorithm. A close reading of some comments suggested that this was a very conservative estimate. The differences between the languages were, according to a two-sample binomial proportion test,<sup>16</sup> 50-60% likely to be significant. Most significant was the suggestion that the Polish comments were more negative than the Ukrainian or Russian ones, but a probability of 60% was too low to accept this conclusion. Rather than revealing difference, the test highlighted how consistent the tone of debate was across languages.

---

<sup>13</sup> <http://wiki.digitalmethods.net/Dmi/ToolWikipediaEditsScraperAndIPLocalizer> (accessed 19 November 2013).

<sup>14</sup> The lesser number of anonymous editors in the case of Ukrainian *Wikipedia* can be explained by the fact that a considerable number of Ukrainian users use Russian *Wikipedia* instead and, consequently, contribute more to the development of the latter project.

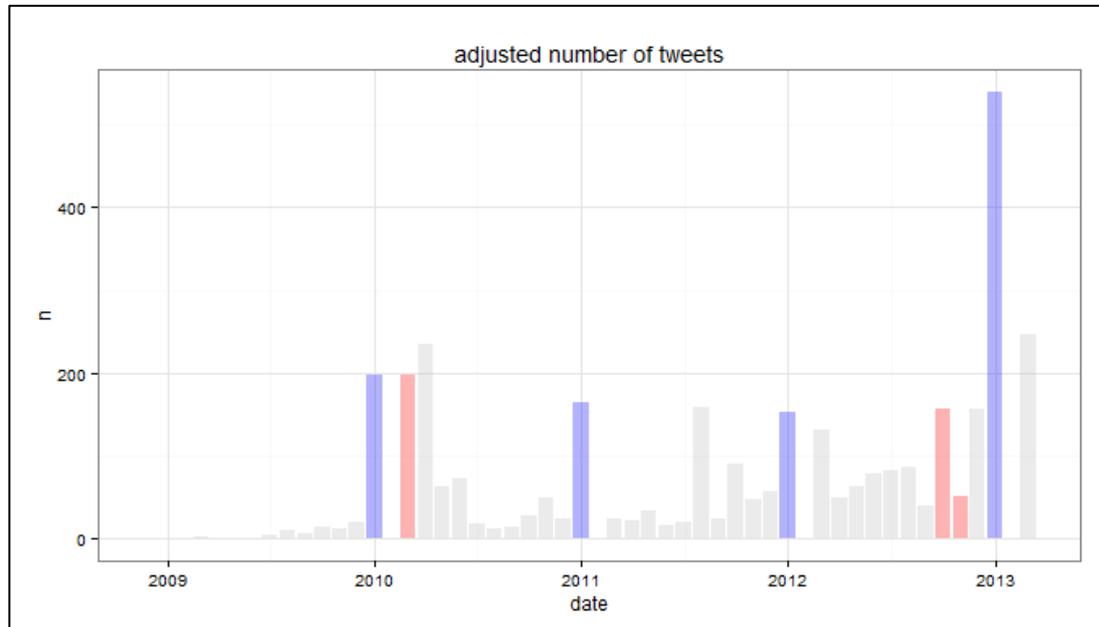
<sup>15</sup> <http://sonetlab.fbk.eu/wikitrip/> (accessed 19 November 2013).

<sup>16</sup> The test is used to evaluate whether two results are equally probable, as in a coin toss. 'Binomial' here means we are testing a binary distinction such as presence or absence, rather than measurement on a scale.

## Bandera on *Twitter*

While *Wikipedia* involves a large number of users, the number of actual contributors is very small. A better indication of perception may be found by accessing *Twitter* data, where individuals can freely express their views on any subject of their choice. Freedom of expression together with the less elitist stance of the platform – *Twitter* accepts all kinds of contributions in contrast to the strict semi-academic guidelines in *Wikipedia* – encourages involvement from a wider audience. Consequently, while *Wikipedia* presents the work of a few contributors to a wide audience, *Twitter* allows for massive engagement both on the producing and receiving ends. *Twitter* data is available online, though it is made available through a restrictive API. Alternatively, the Yandex search engine archive of tweets offers a deeper record of tweets about Bandera than that offered through the *Twitter* API. Although this approach allows time series comparisons, it is problematic in that the archive is partial and made based on nontransparent archiving principles. Yandex appears to collect Ukrainian and Russian tweets equally, but stores few or no Polish tweets. A comparison of Yandex’s record to a complete scrape of *Twitter* for March 2013 revealed that Yandex archived about 50% of all tweets and favoured ‘reputable’ sources. Most of the tweets, roughly two in three, were in Russian.<sup>17</sup>

**Figure 2.** Tweets per month



Source: Rolf Fredheim

<sup>17</sup> This was done using the *cldr* implementation of Google Chrome’s language detection algorithm: <http://cran.r-project.org/web/packages/cldr/index.html>. For details consult the R scripts at <https://github.com/ghowa/bandera-scripts> (accessed 1 December 2014).

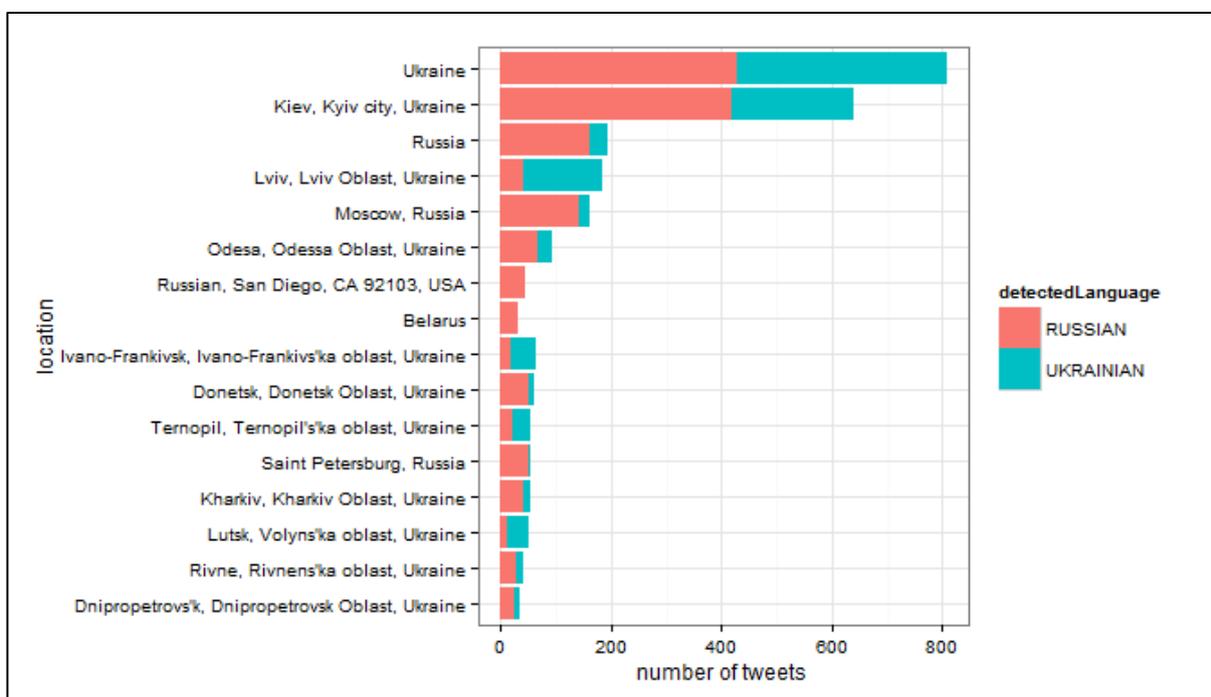
### Distribution over time

Tweets, much more than *Wikipedia* page views, were clustered around predictable key junctures such as elections (red) and annual commemorations (blue). The plot above illustrates that every year in December/January, near the time of Bandera's birthday, the Twittersphere came to life. Both in 2010 and in 2012, election campaigns were accompanied by an increase in *Twitter* activity. The sharp increase in 2013 was probably due to the nature of the archive: some recent 'low-quality' tweets were only briefly included in the data. With the exception of early 2013, which may be an anomaly, overall *Twitter* use appeared reasonably constant.

### Where do the tweets come from?

Out of almost 15,000 tweets 5,500 contained Geolocation data, and of these, over 4,000 were in Ukraine and 1,150 in Russia. These were self-declared locations and liable to misinformation. Locations such as 'the solar system' or 'USSR' are humorous and easily weeded out; more problematic was misinformation, but the sample of 5,500 users was large enough to mean that the number of users misreporting their location as Kyiv rather than Tambov is probably insignificant. While two thirds of tweets were in Russian, roughly four times more tweets originated from Ukraine than from Russia. Within Ukraine, the data revealed an east-west divide with the disproportionate absence of the central and southern regions (with the notable exception of Kyiv), but in no city is one or the other language monolithic:

**Figure 3.** Tweets by language and location



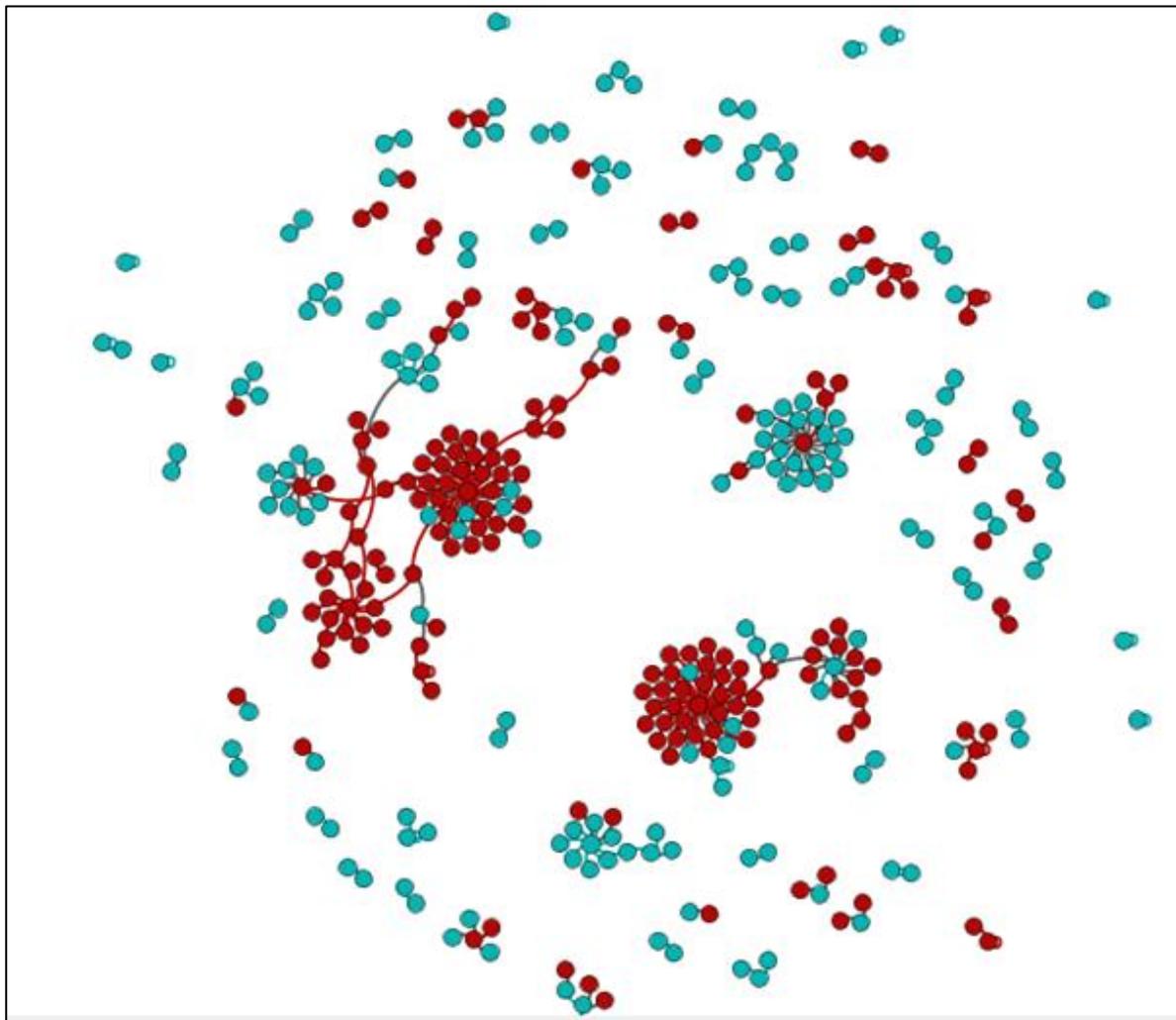
Source: Rolf Fredheim

### *Twitter interactions*

To inspect interactions between *Twitter* users, we took retweets and direct messages to represent interactions between users. We also removed all silent *Twitter* users, that is, users who received tweets, but did not themselves speak about Bandera. We then mapped these connections with the popular software Gephi to reveal patterns of interactions. Although the largest in number, Russian language users had relatively few links with other *Twitter* users interested in Bandera.

The map of the twittersphere reveals considerable isolation, especially for Russian language users (blue):

**Figure 4.** Connections between *Twitter* users (blue: Russian, red: Ukrainian)



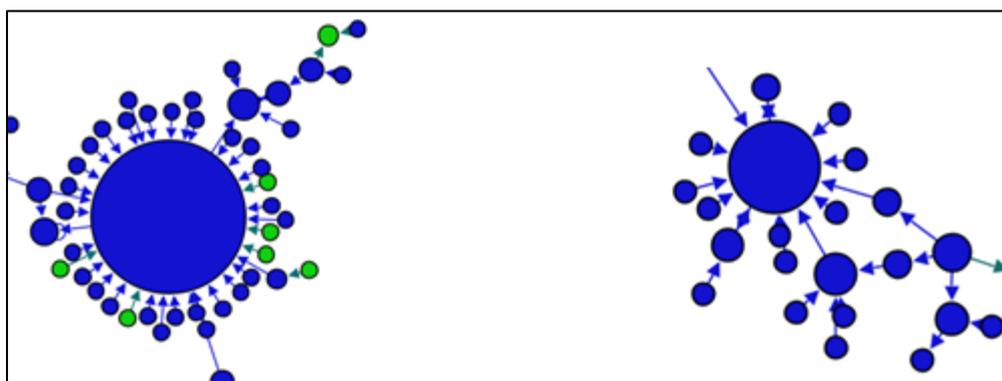
Source: Rolf Fredheim

There were some exceptions, notably the substantial blue cluster in the centre of the visualisation. When using Geolocation rather than detected language to subdivide the tweets, all the major clusters appeared to originate geographically in Ukraine, suggesting that Russians out-

side Ukraine may tweet about Bandera, but do not actively and systematically engage in the main debates.

In the *Twitter* data, there were two main types of content: promotion of other online material, most commonly online newspaper articles or videos and direct interaction between users discussing Bandera. The former category was in much greater evidence in our research, pointing to the dependence of *Twitter* activity on other information sources and, in particular, mainstream media. Such a pattern was already observed in studies concerning the use of *Twitter* during political campaigns (Bruns 2010; Larsson and Moe 2012). The majority of nodes were media sources, which in part reflects Yandex's archiving policy, but also the particular way in which the *Twitter* platform is used. There was a surprising lack of interaction between users. The largest single node was formed around the online newspaper *Ukrains'ka Pravda*. Most of the links with this source were retweets, as users promoted the content for wider circulation.

**Figure 5.** User clusters centred around a news source and around an individual



Source: Rolf Fredheim

There was little or no interaction surrounding Bandera between users who retweeted the same text. Much more interaction was observable around secondary hubs. Valentyn Desiatnyk (@10nyk) – a self-styled ‘Ukrainian bourgeois nationalist’ formed the centre of a purely Ukrainian language group. A second large Ukrainian language cluster was formed around the journalist Olena Bilozirs’ka and the Ukrainian language channel TSN. There was, additionally, a small cluster around Natalia Vitrenko, the leader of the pro-Russian Ukrainian Progressive Socialist Party. This cluster is apparently located in Ukraine, but is entirely Russian speaking and generally hostile:

Stoiat', bandery! Velodirizhabel'naia eskadril'ia PSPU vyletaet spasat' slavianskoe edinstvo! Prekratim #warinrussia, razviazannuiu NATOGestapo! [Stop, Banderas! The Bicycle-Blimp Squadron of the Progressive Socialist Party of Ukraine is flying out to save Slavic unity! Let's stop the #warinrussia, unleashed by the NATOGestapo!] (@Kondratenberg 2011)

This lack of interaction, even amongst the most ardent activists, points to *Twitter* being used more for dissemination of information than for engaging in debate. This observation supports findings from earlier studies, which argued that *Twitter* is more commonly used as a news dissemination resource than as a social networking platform (Kwak et al. 2010).

We also identified a series of key users, who appeared to function as bridges between Ukrainian and Russian speaking clusters. For instance, within the giant cluster (above) the individuals @bihanskyy, @arhe\_dmitriy and @elf\_ira fulfilled such roles. These are individuals, apparently with relatively moderate views and equally comfortable using either language. @bihanskyy lives in L'viv and studied in Chemnitz. He tweets in both Russian and Ukrainian, though Ukrainian appears to be his preferred language. In contrast, @arhe\_dmitriy works for an online bookshop; his preferred language appears to be Russian. Finally, @elf\_ira lives in Kyiv and tweets in both languages. None of these individuals are obviously 'memory warriors', and it appears that, by and large, *Twitter* does not reflect offline conflicts about Stepan Bandera. When there is debate on *Twitter*, it is often about the validity of the sources, but within the data as selected by Yandex there are few hostile exchanges and relatively sparse interaction beyond exchanging links.

### **Bandera on YouTube**

*YouTube* is without doubt the most successful video platform on the Internet. It was founded in 2005 and proved to be an instant success. Just one year later, it was purchased by *Google* (Snickars/Vonderau 2009: 9). According to statistics published by *YouTube* on its website, it has more than 1 billion users each month who watch roughly 6 billion hours/month of video as of 20 November 2013. Every minute, 100 hours of videos are uploaded. Whereas it started as an American website, nowadays 80% of its traffic stems from other countries (YouTube 2013). So although there are national competitors, e.g. *Rutube* in Russia (Gornyxh 2009), *YouTube* nonetheless remains the most popular video platform worldwide. Therefore, it is perfectly suited for our analysis of cross-country, cross-media coverage of Stepan Bandera.

For our basic corpus, we selected both the first 100 videos returning for the search string 'stepan bandera' and those returned for its Cyrillic equivalent 'степан бандера'. This allowed us to identify three principal starting points for a quantitative analysis of these 200 videos. First of all, for most *YouTube* videos there was statistical information available, by means of which we gained a first insight into *YouTube* usage concerning Stepan Bandera. The second step was to scrape all comments from the selected 200 videos and attempt a distant reading by means of statistical analysis. By doing so we were able to learn about the discussions triggered by these videos and language used in these discussions. Third, we implemented tools to 'distant watch' the videos, i.e. to get to know them without actually watching them. Since as of yet, quantitative methods have generally been applied to written texts, this last approach was an innovation for quantitative memory studies.

### Statistics

According to media scholars Pelle Snickars and Patrick Vonderau, *YouTube* is not so much about social networking and community as about data:

YouTube presents videos in conjunction with viewer statistics, not detailed user profiles. As a matter of fact, “users” are by definition reducible to quantified traces of actual usage. With views, clicks, comments and ratings counted, user behavior becomes a byproduct of all the informational transactions taking place on the site, and raw data constantly gets fed back into the YouTube machinery. (Snickars/Vonderau 2009: 16)

Apparently, there is an abundance of statistical data available for each video stored on *YouTube*'s servers. Only the original uploader can access this data directly (Google 2013). A tiny subset of this data, however, is available on each video's *YouTube* page. Although the original uploader can turn off the display of more elaborate statistics, two numbers are always available: overall view count and numbers of comments. Our first step was therefore to scrape this information and create a table of the ten most watched and the ten most discussed Bandera Videos (Tables 1 and 2). Video language and type were determined manually.<sup>18</sup>

**Table 1.** Ten most watched Bandera videos (as of 18 September 2013)

	<i>Title</i>	<i>URL</i>	<i>Lang</i>	<i>Type</i>	<i>Views</i>
1	Tesak o Kavkaztsakh [Tesak about Caucasians]	youtu.be/MsCwHJzRyAk	ru	TV	224687
2	Stepan Bandera. Rassekrechennaia zhizn'. 1 chast' [Stepan Bandera. Declassified Life: Part 1]	youtu.be/4usnbTmSamQ	ru	TV	82269
3	Tribute to Stepan Bandera	youtu.be/pEhJVq5amKs	en	slide	69104
4	Stepan Bandera. Rassekrechennaia zhizn'. 3-8 [Stepan Bandera. Declassified Life: Parts 3-8]	youtu.be/_tInhcfLNd0	ru	TV	26254
5	Marsh 'Stepan Bandera' [March 'Stepan Bandera']	youtu.be/ZMCRkHxOnAU	ru	slide	25555
6	Stepan Bandera na kresach 2/3 [Stepan Bandera in the Kresy 2/3]	youtu.be/JnwuwJH9iTs	pl	TV	23507
7	Between Hitler and Stalin: Ukraine in World War II	youtu.be/mMB5on311Ug	en	TV	23079
8	Stepan Bandera. Rassekrechennaia zhizn'. 6-8 [Stepan Bandera. Declassified Life. 6-8]	youtu.be/SqjpeXVQIr4	ru	TV	17759
9	Stepan Bandera. Rassekrechennaia zhizn'. 1-8 [Stepan Bandera. Declassified Life. 1-8]	youtu.be/9tBxdj_Cu-E	ru	TV	16361
10	Neizvestnye bitvy Rossii 2 (Gimry 1832 g.) [Unknown Battles of Russia 2. (Gimry 1832)]	youtu.be/zspDt4h3h9M	ru	TV	16294

<sup>18</sup> More information about video languages and types can be found in the subsequent section 'Distant Watching'.

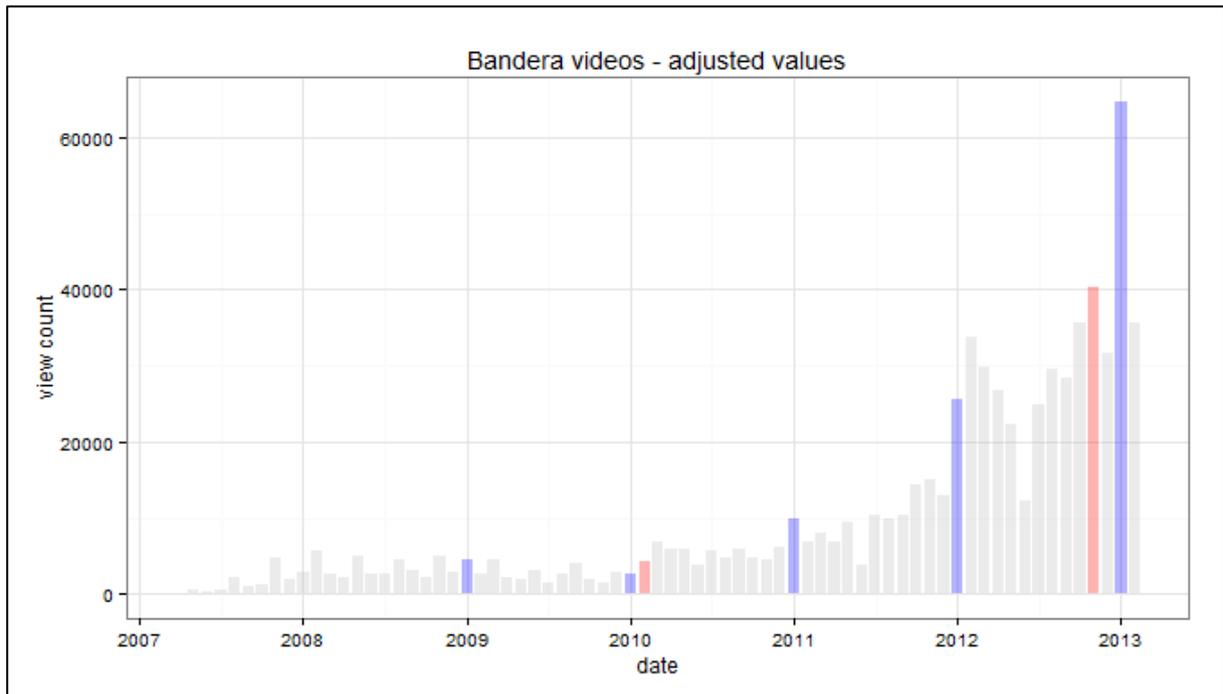
**Table 2.** Ten most discussed Bandera videos (as of 18 September 2013)

	<i>Title</i>	<i>URL</i>	<i>Lang</i>	<i>Type</i>	<i>Comm.</i>
1	Stepan Bandera. Rassekrechennaia zhizn'. 1 [Stepan Bandera. Declassified Life. 1]	youtu.be/4usnbTmSamQ	ru	TV	1088
2	Tesak o Kavkaztsakh [Tesak about Caucasians]	youtu.be/MsCwHJzRyAk	ru	TV	846
3	Fakel'noe shestvie v den' rozhdeniia Bandery ['Torchlight Procession on the Birthday of Bandera']	youtu.be/G2eONYmKXUY	ru	TV	584
4	Stepan Bandera. Rassekrechennaia zhizn'. 3-8 [Stepan Bandera. Declassified Life. 3-8]	youtu.be/_tInhcfLNd0	ru	TV	365
5	Stepan Bandera. Rassekrechennaia zhizn' 2 [Stepan Bandera: Declassified Life. 2]	youtu.be/uwzIt6HyKN8	ru	TV	290
6	Stepan Bandera na kresach 2/3 [Stepan Bandera in the Kresy 2/3]	youtu.be/JnwuwJH9iTs	pl	TV	273
7	Stepan Bandera – heroi? [Stepan Bandera – A Hero?]	youtu.be/d5DzL6RCnTE	ua	TV	238
8	Stepan Bandera heroi Ukrainy [Stepan Bandera – Hero of Ukraine]	youtu.be/VWp2k15lxXQ	ua	slide	228
9	Karat Napalm Grupa – Pod Krwawą Bandera [Karat Napalm Group – Under the Bloody Flag]	youtu.be/2-5hi9QLPJw	pl	slide	223
10	Stepan Bandera na kresach 1/3 [Stepan Bandera in the Kresy 1/3]	youtu.be/ALx69iEZ80Q	pl	TV	187

The most watched video in our corpus, ‘Tesak o Kavkaztsakh’ [‘Tesak about Caucasians’], had nothing to do with Bandera, but instead was focused on immigrants from the Caucasus in Russia and was therefore removed from all subsequent statistics.<sup>19</sup> Generally, both lists are dominated by clips remediated from (Russian) TV, but some user-generated content can be found as well, in the form of various slide shows. After these first insights, we decided to access the more elaborate statistics for those videos whose original uploaders have not deactivated this function. Due to a recent change in the *YouTube* website, the scraping of extended statistics features has become rather complicated. Therefore all of the following statistics are based on the data scraped up until 29 April 2013. All in all, eight of the 200 videos had their statistics disabled. As mentioned above, we also removed the ‘Tesak’ clip. For the remaining 191 videos, we have scraped the view counts and upload dates over time (see Figures 6 and 7).

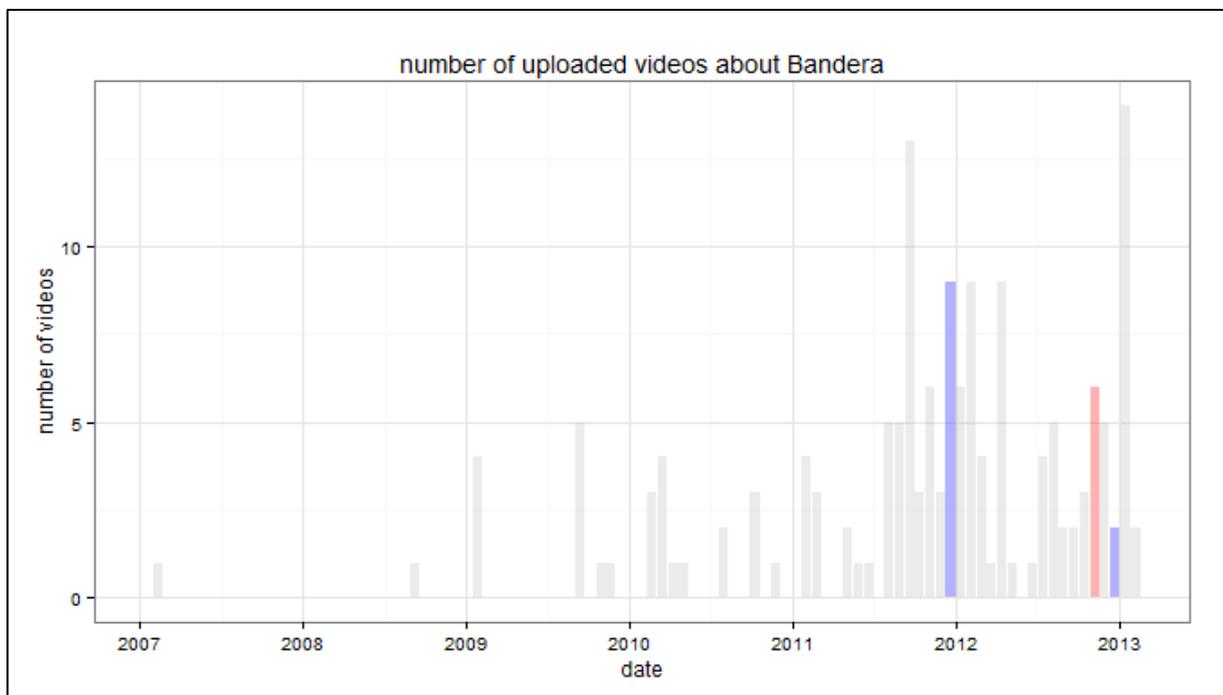
<sup>19</sup> It seems that YouTube included this video in the results because a user named ‘Stepan Bandera’ commented on it: <http://www.youtube.com/watch?v=MsCwHJzRyAk&lc=XxeCC8J7zyp7H3nMUEiS7qGRGZnctmom8-GLxHZ8vU4> (accessed 19 November 2013). This ‘false positive’ underlines the fact that YouTube acts as a type of ‘bottleneck’ (Eijk 2009).

**Figure 6.** Number of Bandera videos watched



Source: Rolf Fredheim

**Figure 7.** Number of Bandera videos uploaded



Source: Rolf Fredheim

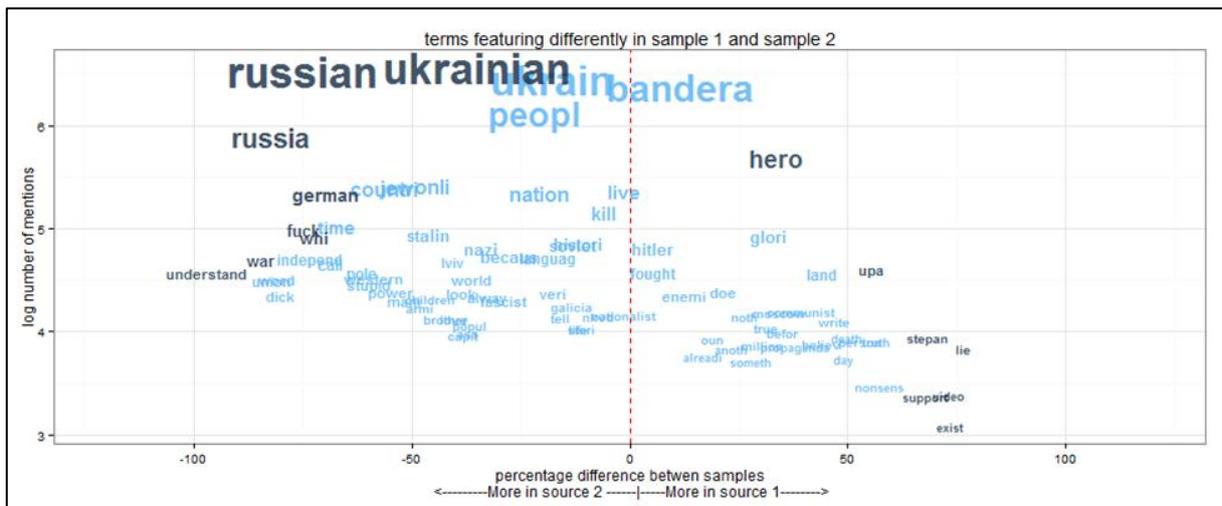
In general, Bandera-related activity on *YouTube* seems to be rising in terms of both videos uploaded and watched. This impression may be deceptive, because the *YouTube* search-ranking will tend to prioritise more recent content, which may consequently be deemed more relevant. Nonetheless, some distinct patterns emerge from the plots: Bandera's birthday on 1 January regularly caused some extra activity, especially in 2009 – his 100th birthday – and onwards. One exception to this rule is 2010, where the proclamation of Bandera as a Ukrainian national hero on 22 January completely overshadows Bandera's birthday. This political decision also seems to have had a certain impact on overall *YouTube* usage: since 2010, Bandera videos have been uploaded on a more regular basis. The parliamentary elections in Ukraine on 28 October 2012 also coincide with an activity spike. It seems as if *YouTube* use has a more lasting memory potential; although the largest number of videos was uploaded in 2012, interest was much higher in 2013. Consequently, much of the interest in Bandera in 2013 was generated by old content – new material had often not yet been added. This demonstrates that since 2012, *YouTube* has been actively used for content about Bandera and that dissemination of the content has become more effective. This observation points to a steady growth of interest in Bandera's personality, coinciding with the memory politics of former President Viktor Yushchenko. Online activity may serve as an important indicator of the ongoing memory processes in Ukrainian society – even while on the official level the nationalist narrative of the Second World War becomes to a certain degree marginalised (Solod'ko 2013).

### *Comments*

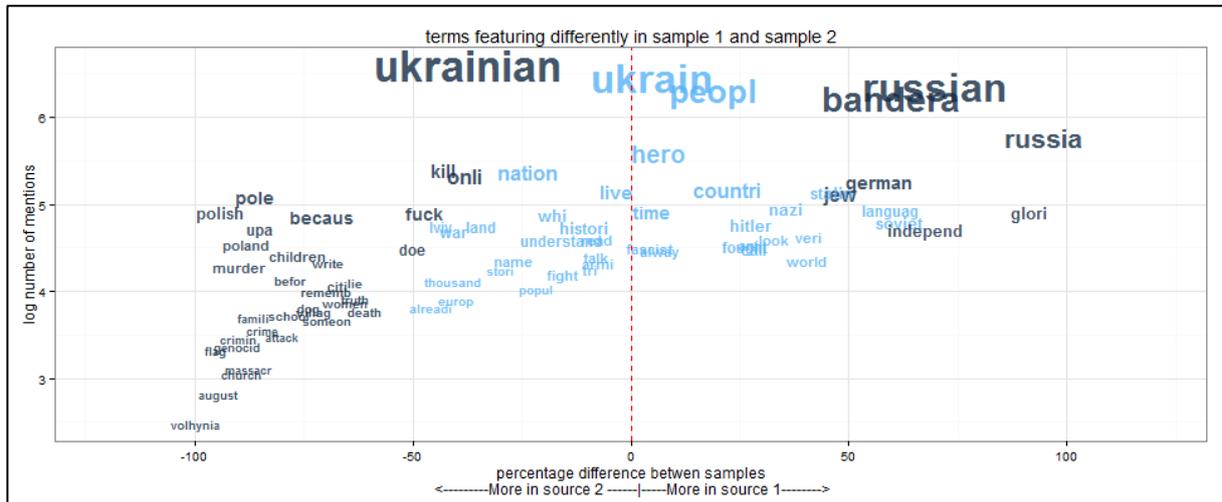
*YouTube* activity is not limited to watching and uploading videos. Another vital part is commenting on them. Hence, we scraped the comments for the 159 videos in our data set for which commenting was enabled, ran language detection and split these into language groups. This corpus represented the aggregate data of all Ukrainian, Russian, Polish and English comments from a large (though not exhaustive) sample of *YouTube* videos. We then used a stemming algorithm and a stop word list to reduce duplicate terms and remove words unlikely to be significant. Finally, each of the language pairs were contrasted and the differences between the two visualised (see Figures 8-10). In these plots, words in dark blue are overrepresented to a statistically significant degree (probability calculated at the 0.95 confidence interval). The visualisations make clear which memory models and narratives are most common across national groups.

The main differences between Ukrainian and Russian language comments (Figure 8) can be reduced to a Russian-Ukrainian dichotomy: 'Russian' and 'Ukrainian' are overrepresented in Russian texts, along with an emphasis on the problematic history of relations between Bandera and the Third Reich ('German'). On the Ukrainian side there is a stronger focus on UPA and Bandera as a hero ('hero', 'UPA'). Most interestingly, the sparsity of significant differences shows that most ideas are present in both language groups. When contrasting either Polish and Ukrainian or Polish and Russian comments, no such balance can be found.

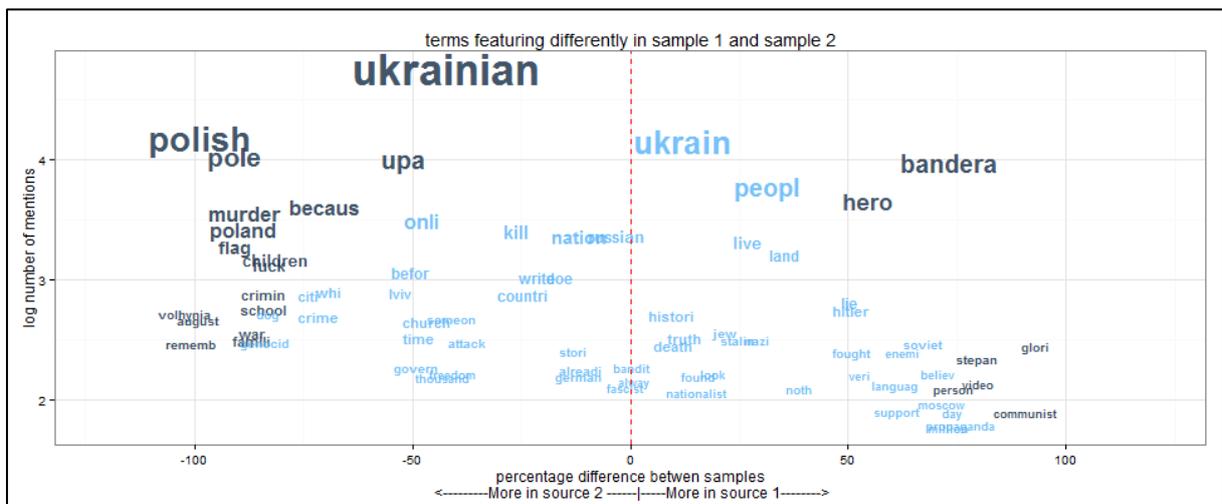
**Figure 8.** Russian versus Ukrainian comments



**Figure 9.** Polish versus Ukrainian comments



**Figure 10.** Polish versus Russian comments



Source Fig. 8-10: Rolf Fredheim

Figure 9 illustrates Polish-Ukrainian differences. In contrast to the Russian-Ukrainian case, here some associations are strongly overrepresented – especially in the Polish comments (from 70 to 100%). The Polish narrative is clearly centred around the Volhynia massacre, which is closely associated with such terms as ‘criminal’, ‘children’ and ‘murder’. Both Nazi and communist references such as ‘Hitler’, ‘Nazi’, ‘Jew’, ‘Stalin’, ‘Soviet’, etc. feature more in Ukrainian comments. Stepan Bandera figures more strongly on the Ukrainian side, while UPA is a major actor in Polish comments. Such terms as ‘Polish’, ‘Poland’ and ‘Pole’ are almost exclusively present in Polish texts, demonstrating the degree to which the Ukrainian commentators fail to acknowledge the Polish narrative about the past and/or the role of Bandera in Polish history.

One might expect a degree of similarity between the Russian and Polish arguments about Bandera, but in fact, the Polish and Russian texts exhibit the starkest differences (Figure 10). Again, the argument about UPA, Volhynia and Poles is virtually absent on the Russian side, which is surprising given that the Russian comments are so numerous (85,000 words). The Russian texts overemphasise Bandera’s role in Ukrainian independence, while both Nazi and Communist connections are overrepresented. The hypothesis that Russian texts would exhibit a greater focus on Bandera as a Nazi collaborator is confirmed, but the hypothesis that Poles would emphasise the Communist connection remains unsubstantiated.

### *‘Distant watching’*

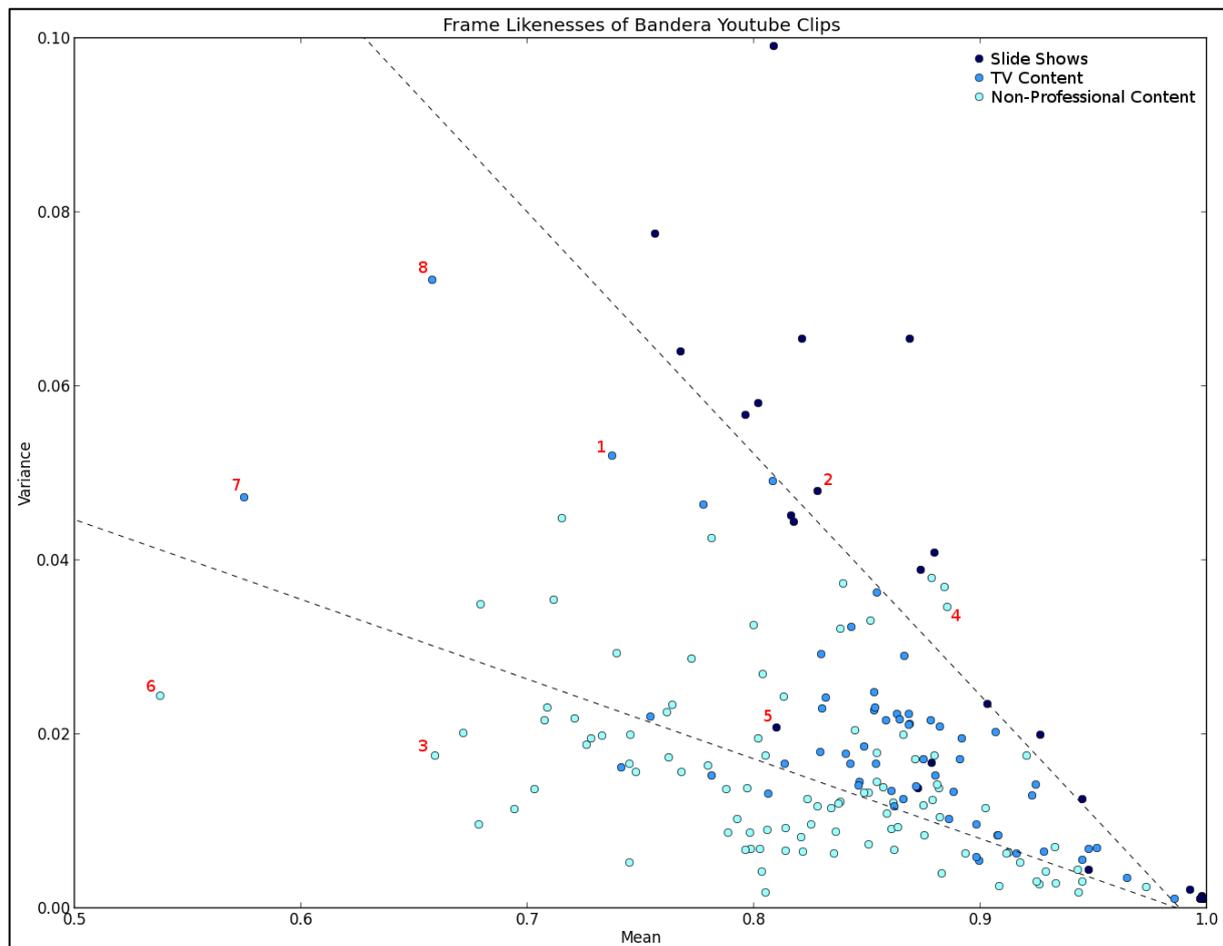
Having learned the basic dynamics of uploading and viewing the videos as well as commenting on them, we need to find out more about the videos themselves. As they last for dozens of hours, it seems reasonable to automatically categorise them in a meaningful way – even more so as the ‘distant watching’ scripts can now be reused for other research questions related to *YouTube*. Our approach towards categorizing *YouTube* videos was to perform a rudimentary genre analysis on them. In our top ten lists (Tables 1 and 2) we can identify two basic genres: remediated TV content such as the Russian documentary ‘Tainy razvedki’ [Secrets of the Secret Service] (RusOsetia 8.3.2013, marked with ‘1’ in Figures 11 and 12) and user-generated videos. User-generated content usually consists of slide shows with music, where each image is shown for several seconds (e.g. CiemnaStrefa 26.10.2012, marked with ‘2’). Thus, the differences between the individual frames of the video are rather small, with abrupt changes in between. On the other hand, the differences between individual frames are higher for TV content, but altogether more continuous. By means of measuring the distances between individual frames, we implement automated genre detection.<sup>20</sup> This automatic attribution helps us to further study the details and dynamics of the different *YouTube* genres we have identified.

We extracted the videos’ frames with a frequency of one frame per second. Thus, from a one-minute video we received 60 individual frames. We then computed the differences for subsequent frames and plotted them using the mean value of the frame differences on the x-

<sup>20</sup> Automatic genre detection for online content was proposed by Mehler, Sharoff and Santini (2010), although they focussed mostly on written texts.

axis and their variance on the y-axis. Afterwards we manually defined the video genres. In most cases, video title and description held enough information to quickly assess the genre, in other cases it was necessary to actually take a glimpse of the video. During this process a third video genre could be detected: non-professional videos filmed using consumer cameras or smartphones (e.g. Stepan Bandera 25.8.2011, marked with '3'). The resulting plot can be found in Figure 11.

**Figure 11.** Frame differences of the 200 *YouTube* clips, coloured according to their genre



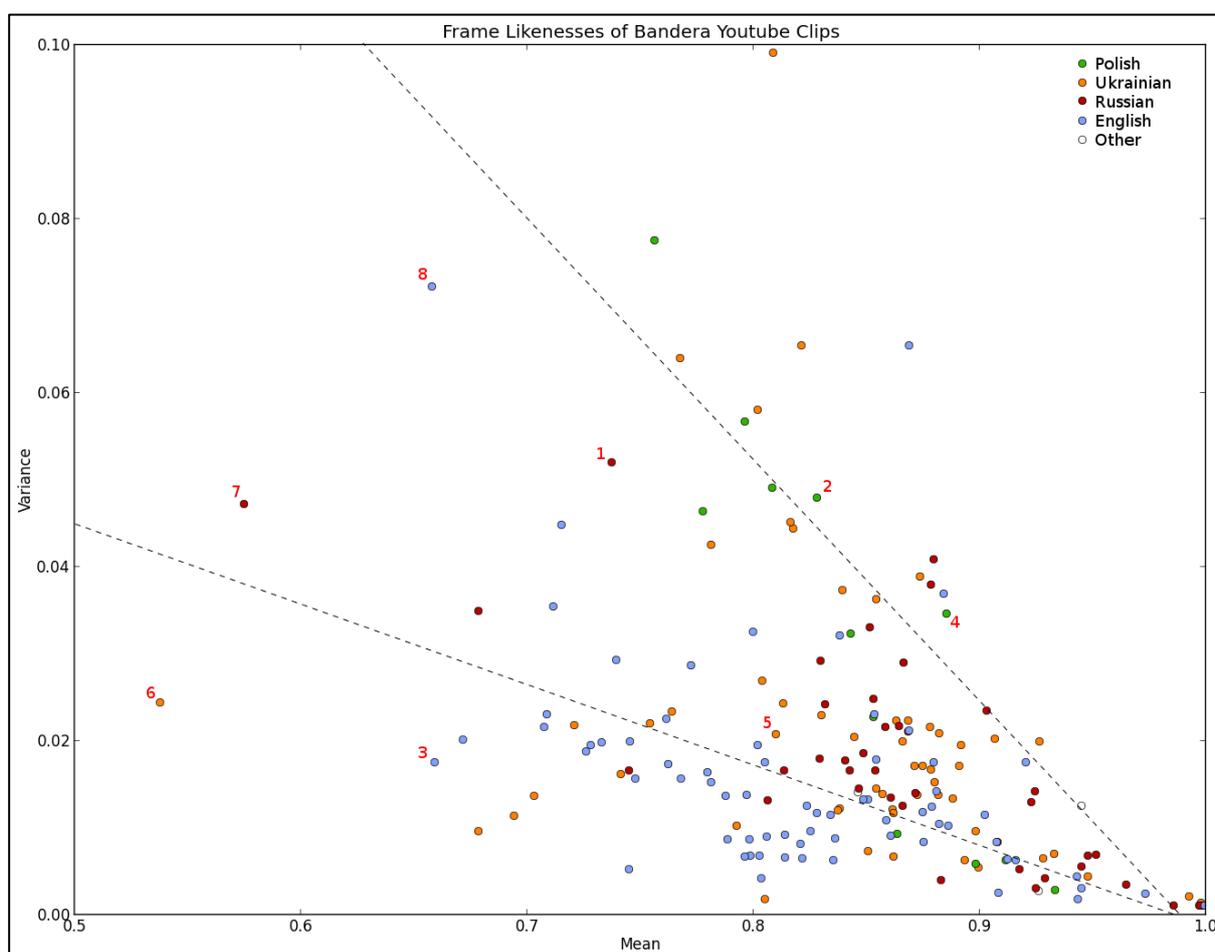
Source: Gernot Howanitz

Although the three genres overlap, a certain order is recognizable. Slide shows are mostly located in the top right corner of the plot. Some very static life action videos are also to be found here (warsztatyTV 8.11.2012, marked with '4'). On the other end of the spectrum are non-professional videos (bottom left). They are defined by constant slight frame changes with few to no cuts. As soon as there are cuts, those private videos can get mixed up with TV content, mostly located in the center of the plot. A few slide shows are also to be found in this region, especially when visual effects such as rotation and panning are used (rurytch 28.1.2010, marked with '5'). In order to ease interpretation, we manually insert two helper lines into the plot to separate the three main genres. The graphic now allows us to select indi-

vidual videos for a closer, in-depth analysis. This partitioning is not specific to Bandera videos but can be applied to other research questions as well. Thus, manual attribution of videos to their respective genres is rendered unnecessary for future projects.

Next we combined this crude genre distinction with language detection. The language of a video was determined based on its title and description retrieved from *YouTube*. In roughly 75% of the cases, this approach worked without problems; the languages of the remaining 25% had to be determined manually. The result of this effort is displayed in Figure 12, again with the genre borders manually inserted.

**Figure 12.** Frame differences of 200 *YouTube* clips, coloured according to their language



Source: Gernot Howanitz

According to this plot, slide shows seem to be mostly in Ukrainian and Polish. This genre is the most basic way to create a video. TV content is mainly Russian and Ukrainian, with mostly documentaries (RusOsetia 8.3.2013, '1') and a few news clips (RT Russian 1.1.2013, '7'; AFP 2.1.2013, '8'). English predominates in non-professional videos, which can be attributed to a large Ukrainian community that is based in San Francisco and documents its activities under the *YouTube* username 'Stepan Bandera' (e.g. Stepan Bandera 25.8.2011, '3').

This automatic assessment of video genres did not tell us much about content, although it could to a certain extent identify similar videos. For example, the three leftmost videos, marked with ‘6’ through ‘8’, all covered a torchlight procession in Kyiv on 1 January 2013. Video 6 was an example of a non-professional account of the event (Aronets’ 1.1.2013), whereas video 7 and 8 were news clips from *Russia Today* (RT Russian 1.1.2013, ‘7’) and *Agence France-Presse* (AFP 2.1.2013, ‘8’).<sup>21</sup> This example shows how the graphics produced by the ‘distant watching’ scripts point our attention to videos we would not have watched otherwise. Moreover, they allow us to trace the relationships between individual videos.

## Analysis

Throughout this study, we observed that the majority of the Bandera material was produced in the Russian language, but geographically in Ukraine. This may be explained by Bandera’s image in Ukraine, where he is perceived less as a historical personality and more as the source of conflicting historical narratives in the Russian-speaking East and Ukrainian-speaking West (Portnov 2014). Three times as many *YouTube* comments and twice as many tweets were in Russian, while the Russian language *Wikipedia* page about Bandera was three times as long as the Ukrainian one and the underlying discussion was six times as long. Does this suggest that the Russians care most about Bandera, perhaps because he figures as an easily identifiable villain? Any such interpretation is rapidly refuted by examining the origin of online interactions. Across all three media we observed that activity was strongly concentrated in Ukraine. Does this suggest that Bandera is contested as part of a national Ukrainian debate, one that moreover is conducted primarily in Russian?

This appears to be true, but only to a degree. When we examined the expressed attitudes towards Bandera, we found surprisingly little difference, not just between Russian and Ukrainian, but all the languages within the online platforms. In a series of statistical tests, we recognised virtually no difference between the proportion of negative words used in different language versions of the *Wikipedia* article about Bandera, the discussion page, or indeed on *YouTube*. Only on *Twitter* did Russian speakers use significantly more negative language than did Ukrainians. Interestingly, however, we were able to detect sharp divisions between different social media platforms: *Wikipedia* discussions were the most moderate while *YouTube* comments emerged as twice as negative. These patterns were consistent also for Polish and English.

In this instance it appears that genre and different usage patterns were remarkably consistent across languages. *YouTube* comments were equally marked by diatribes and vulgarities for every language examined. *Twitter*, in contrast, was used to disseminate information.

---

<sup>21</sup> One further possible step to automatically identify video content would be to use image recognition algorithms to locate common images of Stepan Bandera or symbols such as the Nazi swastika, the Soviet hammer and sickle, and the Ukrainian coat of arms within the extracted video frames. Thus, one would know if a video is indeed about Bandera, and could also identify the different historical contexts in which he is presented. Unfortunately, image recognition is a very complicated task that would have gone beyond the scope of this article.

Most *Wikipedia* content was neutral in tone, but pointed to other resources that might be notably less so.

While expressions appear conditioned more by the social platforms employed by users, the large quantity of Russian language material points to Russian material as a site of conflict. This is not surprising, given that, within the former states of the Soviet Union, Russian often works as a lingua franca and that Ukrainian speakers are more willing to use Russian than Russian speakers are to use Ukrainian. Assuming that conflict is more likely in fora available to people holding opposing views, we expected to see a higher incidence of othering, or ‘us-them’ polarisation in Russian texts.<sup>22</sup> This, however, was not the case. When exploring *YouTube* comments and comparing Ukrainian and Russian language statements, we saw that actually, those written in Ukrainian more often used strong othering language. In this regard, it is worth noting the role played by English language resources, which are also accessible equally to all language groups. Across the board, English content was the most negative; this was especially true for *YouTube*, where English was also the most polarised. Though Russian and English both figured as linguae francae, the former proved the most moderate, while the latter saw the greatest hostility and polarisation. What explains this difference?

English *YouTube* comments were probably the most aggressive because this was the main site where speakers of different languages interact. In addition, English language resources appeared to be the only sites where the Polish and Ukrainian narratives met. While comments on English language content were the most aggressive, it is Polish and to a lesser degree Ukrainian content that was the most antagonistic. In this regard, our findings contradicted Volodymyr Kulyk’s, who concluded that ‘[...] the inherently transnational Internet communication undermines the national framework for the production of identity’ (Kulyk 2013: 79). The plot of genre distribution in *YouTube* videos showed that most user-generated content about Bandera was in the Polish or Ukrainian language, while Russian videos tended to be clips from television, reposted to *YouTube*.

Ukrainian-Russian debates tended to revolve around interpretations of Bandera’s historical role, while Polish debates focused on the alleged genocide conducted by Ukrainian nationalists, associated with Bandera, against the Polish population in former Polish territories. Not only did Russians and Ukrainians rarely engage with this Polish argument online, it was also virtually the only aspect of Bandera debated in Polish comments. The most popular video was a TVP documentary about hardships faced by Poles living in eastern Ukraine and Stepan Bandera’s role in the Volhynia massacre, while the second most popular video, released on 26 October 2012, immediately prior to the Ukrainian parliamentary elections and promoted on Facebook, was a rap called ‘Pod Krwawą Bandera’ [‘Under the Bloody Flag’]. It tells the story of Bandera and the Volhynia massacre in a slideshow format, featuring provocative imagery of dead children and highly inflammatory lyrics.

---

<sup>22</sup> For a discussion of the underlying rationale and technical explanation, see Fredheim (2013b).

## Conclusions

### *Patterns of (non-)interaction*

One might expect that Poland would figure as a distinct case in the Stepan Bandera case study, but the extent to which Polish voices failed to interact with Ukrainian and Russian ones online is striking. Though evidence from the *Twitter* data was lacking in this regard, the *Wikipedia* edits revealed that while the Ukrainian and Russian pages contained a large number of international edits and contributions to discussions, the Polish page was edited almost exclusively by users writing in Polish and located within Poland. Associated with this was a sharp difference in focus between the Polish debates and the Russian-Ukrainian debates, where the former focused almost exclusively on the massacre of Poles while the latter overlooked it. The only exception was the small number of English language videos on *YouTube*, which saw the most aggressive and polarising rhetoric.

In general, location appeared to trump language: that is, those using Russian in Ukraine were much more engaged in debates about Bandera than those using Russian in Russia. The Russian *Wikipedia* entry about Bandera, for instance, was edited mainly by users located in Ukraine. On *Twitter*, Russian users in general, but especially outside Ukraine, were relatively isolated. On *YouTube*, however, Russian language content was the most popular and the site of most debate, generally centred on interpretations offered by documentaries, or assessments of the Bandera marches.

Sentiment analysis of user responses showed greater differences between genres than between language groups. However, in terms of *YouTube* video content, Russian language resources were the least overtly hostile, while Polish resources were the most hostile, even though we hypothesised the opposite.

*Twitter* activity was sparked most by both Bandera's birthday and the Ukrainian elections. For *Wikipedia*, this pattern was similar, but with a certain amount of continuous baseline activity. *YouTube* usage was somewhat different: cyclical events were of less importance, but dates were sometimes targeted by video creators (e.g. 'Pod Krwawą Bandera'). While *Wikipedia* interest tended to be stable, with less extreme peaks of interest in recent years, the number of *YouTube* views seemed to grow exponentially. *Twitter* usage also appeared to be relatively stable. The data therefore suggests that video content is ever more frequently used in the debate about Bandera.

### *The value and pitfalls of distant reading and watching*

The idea of using statistical methods to get an overview of online activity about Stepan Bandera proved to be tremendously helpful, but also extremely labour-intensive. We were able to adapt various 'distant' techniques in order to work with all of the three platforms in question: *Wikipedia*, *Twitter* and *YouTube*. Of course, the difficulty of obtaining statistics varied greatly. The easiest approach was to use statistics generated by the websites themselves: view counts, upload counts, page views, etc. However, the platforms often change the amount of data accessible and also limit access to this data through various means.

Compared to simply scraping statistical data, it was more challenging to ‘distant read’ texts. We had to overcome various problems, e.g. translating the texts to make them comparable or filtering common words that only add noise. Fortunately, many recent research projects focus on ‘distant reading’ texts, and so they can be used with only slight adaptation. In general, the ‘distant reading’ strategy is not as dependent on a platform’s ‘good will’: if a text is accessible on the Internet, it can be analysed.

It should not be a surprise that ‘distant watching’ proved to be the most difficult. Videos are very complex to handle and it takes significant computation time and storage space to extract statistical information from them. The most productive approach was to calculate image likeness. This way, sharp cuts can be identified that point to genre. We envisioned an actual assessment of video content by means of applying image detection algorithms on individual frames. Although identifying specific symbols or photographs within videos worked occasionally, error rates were too high for this approach to be useful in this study. Nonetheless, videos are an important means of online activity and creativity and deserve attention from the quantitative side, regardless of difficulty. Our ‘distant watching’ approach can be regarded as a first step in this direction.

While this article concentrates on the technical aspects of studying cultural memory online and glosses over ‘classic’ approaches to researching cultural memory, we hope we have shown which statistical approaches towards online platforms might be fruitful, which tools exist and how they can be combined to form the lenses of quantitative memory studies. Without doubt, further research is necessary to fully reconcile traditional theories of cultural memory with quantitative methods. We hope that this study and our data will contribute to the foundation of such a practice.

## References

- AFP (2013, 2 January). ‘Nationalists March with Torches in Kiev’. *YouTube*, <http://www.youtube.com/watch?v=N7G1-igqhQ8> (accessed 10 September 2013).
- Aronets’, Oleksandr (2013, 1 January) ‘Den’ Narodzhennia Bandery 2013. Kyiv’ / ‘Birthday of Bandera 2013. Kyiv’. *YouTube*, [http://www.youtube.com/watch?v=CXH\\_xOfCLQA](http://www.youtube.com/watch?v=CXH_xOfCLQA) (accessed 10 September 2013).
- Balyns’kyi, Ihor, Tarik Syril Amar and Iaroslav Hrytsak (2010). *Strasti za Banderoiu / Passions for Bandera*. L’viv: Grani-T.
- Bruns, Axel (2010). ‘Key Events in Australian (Micro-)Blogging during 2010: Paper Presented at the ECREA 2010 Conference, Hamburg’. <http://www.slideshare.net/Snurb/key-events-in-australian-microblogging-during-2010> (accessed 14 July 2014).
- CiemnaStrefa (2012, 26 October). ‘Karat Napalm Grupa – Pod Krwawą Bandera’ / ‘Karat Napalm Group – Under the Bloody Flag’. *YouTube*, <http://www.youtube.com/watch?v=2-5hi9QLPJw> (accessed 10 September 2013).
- Dounaevsky, Hélène (2013). ‘Building Wiki-History: Between Consensus and Edit Warring’ in *Memory, Conflict and New Media: Web Wars in Post-Socialist States*, edited by Ellen Rutten, Julie Fedor and Vera Zvereva. London: Routledge, 130-142.

- Eijk, Nico van (2009). 'Search Engines, the New Bottleneck for Content Access', in *Telecommunication Markets: Drivers and Impediments*, edited by Peter Curwen, Julius Hauscap and Brigitte Preissl. Heidelberg: Physica, 141-156.
- Etkind, Alexander (2010). 'Mapping Memory Events in the East European Space', in *East European Memory Studies*, 1: 4-5, <http://www.memoryatwar.org/enewsletter%20oct%202010%20final.pdf> (accessed 23 February 2014).
- Etkind, Alexander (2013). 'Mourning and Melancholia in Putin's Russia: An Essay in Mnemonics' in *Memory, Conflict and New Media: Web Wars in Post-Socialist States*, edited by Ellen Rutten, Julie Fedor and Vera Zvereva. London: Routledge, 32-47.
- Fredheim, Rolf (2013a). 'Detecting Memory Events'. *East European Memory Studies*, 14: 9-14, [www.memoryatwar.org/enewsletter-May-2013.pdf](http://www.memoryatwar.org/enewsletter-May-2013.pdf) (accessed 23 February 2014).
- Fredheim, Rolf (2013b). 'Quantifying Polarisation in Media Coverage of the 2011-12 Protests in Russia'. *Digital Icons: Studies in Russian, Eurasian and Central European New Media*. 9, 27-49, <http://www.digitalicons.org/issue09/rolf-fredheim/> (accessed 14 July 2014).
- Gornykh, Andrei (2009). 'From YouTube to RuTube, or, How I Learned to Stop Worrying and Love All Tubes' in *The YouTube Reader*, edited by Pelle Snickars and Patrick Vonderau. Stockholm: National Library of Sweden, 441-455.
- Hordasevych, Halina. (2001). *Stepan Bandera: liudyna i mif / Stepan Bandera: The Man and the Myth*. L'viv: Piramida.
- Hu, Mingqing and Bing Liu (2004). 'Mining and Summarizing Customer Reviews'. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177. [www.cs.uic.edu/~liub/.../kdd04-revSummary.pdf](http://www.cs.uic.edu/~liub/.../kdd04-revSummary.pdf) (accessed 14 July 2014).
- @Kondratenberg (2011, 11 April). 'Stoiat', bandery!'. *Twitter*. <https://twitter.com/Kondratenberg/status/57438707774537730> (accessed 1 December 2014).
- Kulyk, Volodymyr (2013). 'War of Memories in the Ukrainian Media: Diversity of Identities, Political Confrontation, and Production Technologies' in *Memory, Conflict and New Media: Web Wars in Post-Socialist States*, edited by Ellen Rutten, Julie Fedor and Vera Zvereva. London: Routledge, 63-81.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). 'What is Twitter, a Social Network or a News Media? Paper presented at the 19th International Conference on World Wide Web'. [an.kaist.ac.kr/traces/WWW2010.html](http://an.kaist.ac.kr/traces/WWW2010.html) (accessed 14 July 2014).
- Larsson, A. O., & Moe, H. (2012). 'Studying Political Microblogging: Twitter Users in the 2010 Swedish Election Campaign'. *New Media & Society*, 14 (5): 729-747.
- Massa, Paolo and Frederico Scrinzi (2013). 'Manypedia: Comparing Language Points of View of Wikipedia communities'. *First Monday*, 18 (1). <http://firstmonday.org/ojs/index.php/fm/article/view/3939/3382> (accessed 14 July 2014).
- Marples, David (2006). 'Stepan Bandera: The Resurrection of a Ukrainian National Hero'. *Europe-Asia Studies*, 58 (4): 555-566.
- Marples, David (2007). *Heroes and Villains*. Budapest: Central European University Press.
- Marres, Noortje and Esther Weltevrede. (2013). 'Scraping the Social? Issues in Live Social Research'. *Journal of Cultural Economy*, 6 (3): 313-335.

- Mehler, Alexander, Serge Sharoff and Marina Santini (eds.) (2010). *Genres on the Web: Computational Models and Empirical Studies*, Dordrecht: Springer 2010. (Text, Speech and Language Technology 42).
- Moretti, Franco (2000) 'Conjectures on World Literature', in *New Left Review* 1: 54-68. [newleftreview.org/article/download\\_pdf?id=2094](http://newleftreview.org/article/download_pdf?id=2094) (accessed 23 February 2014).
- Paulsen, Martin (2013). '#Holodomor: Twitter and Public Discourse in Ukraine', in *Memory, Conflict and New Media: Web Wars in Post-Socialist States*, edited by Ellen Rutten, Julie Fedor and Vera Zvereva. London: Routledge, 82-98.
- Portnov, Andrej (2014). 'Bandera: Lichnost' i Mif' / 'Bandera – Personality and Myth'. <http://urokiistorii.ru/blogs/andrei-portnov/52103> (accessed 14 July 2014).
- Prus, Edward (2004). *Stepan Bandera, 1909-1959: symbol zbrodni i okrucieństwa / Stepan Bandera, 1909-1959: The Symbol of Crime and Atrocity*. Wrocław: Nortom.
- RT Russian (2013, 1 January) 'Fakel'noe shestvie v den' rozhdeniia Bandery' / 'Torchlight Procession on the Birthday of Bandera'. *YouTube*, <http://www.youtube.com/watch?v=G2eONYmKXUY> (accessed 10 September 2013).
- rurytch (2010, 28 January). 'Stepan Bandera-HEROI UKRAIINY-m.mpg' / 'Stepan Bandera-HERO OF UKRAINE-m.mpg'. *YouTube*, <http://www.youtube.com/watch?v=MJ3lnm-3pM8> (accessed 10 September 2013).
- RusOsetia (2013, 8 March). 'Tainy razvedki. Likvidatsiia Stepana Bandery' / 'Secrets of the Secret Service: Liquidation of Stepan Bandera'. *YouTube*, <http://www.youtube.com/watch?v=MmzHJ4S3Tsg> (accessed 10 September 2013).
- Solod'ko, Pavlo (2013). 'V Ukraini nemaie derzhavnoi istorichnoi polityky' / 'No State Historical Politics in the Ukraine'. <http://www.istpravda.com.ua/articles/51dedda02875c/> (accessed 2 July 2014).
- Snickars, Pelle and Patrick Vonderau (2009). 'Introduction' in *The YouTube Reader*, edited by Pelle Snickars and Patrick Vonderau. Stockholm: National Library of Sweden, 9-21.
- Stepan Bandera (2011, 25 August) 'Ukrainian Festival – 2011 in Davis, California, part 9'. *YouTube*, <http://www.youtube.com/watch?v=LmHuHv8wd2g> (accessed 10 September 2013).
- Voss, Jakob (2006). 'Collaborative Thesaurus Tagging the Wikipedia way. Wikimetrics'. <http://arxiv.org/pdf/cs.IR/0604036> (accessed 8 June 2013).
- warsztatyTV (2012, 8 November). 'Portnow Stepan Bandera'. *YouTube*, <http://www.youtube.com/watch?v=tHYycKM4bTc> (accessed 10 September 2013).
- Wikipedia 2013. 'Categorization'. <http://en.wikipedia.org/wiki/Wikipedia:Categorization> (accessed 23 July 2014).
- YouTube (2013). 'Statistics'. *YouTube*, <http://www.youtube.com/yt/press/statistics.html> (accessed 13 September 2013).

ROLF FREDHEIM studied for his MPhil at St Antony's College, Oxford, where his thesis explored the role Katyn played in contemporary Polish political discourse. As a PhD student in Slavonic Studies at the University of Cambridge he was a member of the Memory at War team. At present he is a Research Fellow working on the Conspiracy and Democracy Project at the University of Cambridge, CRASSH. His work focuses on digital approaches to post-

Soviet memory as manifest in the Russian blogosphere and media. He has worked on developing a set of tools for quantifying and presenting trends in large datasets, focusing in particular on memory of the 1990s and events such as the August Putsch. [Ref38@Cam.ac.uk]

GERNOT HOWANITZ studied Russian Literature and Applied Computer Science in Salzburg, Moscow and Prague. He received his MA in Russian Literature in 2011 and his MS in Computer Science in 2013. Currently he is a PhD student and part-time lecturer at the Chair of Slavic Literatures and Cultures at the University of Passau. His research interests include contemporary Russian, Polish and Czech literature, internet culture, computer games and digital humanities. His PhD thesis is supported by a grant from the Austrian Academy of Sciences and focuses on (auto-)biographical strategies currently emerging on the internet in order to analyse different approaches towards the online self-representation of Russian writers. [Gernot.Howanitz@Uni-Passau.de]

MYKOLA MAKHORTYKH studied Archaeology and Euroculture in Kyiv, Cracow and Göttingen. He received his MA in Archaeology in 2009 and his MA in Euroculture in 2011. Currently he is a PhD student at the Institute of Culture and History at the University of Amsterdam. His research looks at World-War-II memory in Ukraine and how it is affected by the processes of de-Sovietisation, nationalisation and digitisation that the country is currently undergoing. In his doctoral thesis he unravels the digital discourse surrounding two milestones in Ukrainian war memory – the capture of Lviv and of Kyiv – in a selection of several social media platforms. [M.Makhortykh@Uva.nl]