

Distant Watching. Ein quantitativer Zugang zu YouTube-Videos

Mag. Dipl.-Ing. Gernot Howanitz

Stipendiat der Österreichischen Akademie der Wissenschaften (DOC)
am Lehrstuhl für Slavische Literaturen und Kulturen, Universität Passau

Zusammenfassung

Um virale Videos auf *YouTube* untersuchen zu können, ist ein quantitativer Zugang unumgänglich. Die Literatur- und Kulturwissenschaften – insbesondere die Slavistik – beschränken sich allerdings größtenteils auf textbasierte Verfahren. Das hier vorgestellte ‚distant watching‘ kann die Schranken des Textes überwinden, weil es erlaubt, eine große Anzahl an Videos mit einem Blick zu erfassen und Rückschlüsse auf deren grundlegende kreative Mechanismen zu ziehen.

1. Überblick

Ein nicht unbedeutender Teil der Internetkultur besteht aus Bildern und viralen Videos, die kopiert, verändert und weiterverbreitet werden und damit zum ‚Mem‘ werden (Dawkins 2013; Shifman 2013). Die schiere Masse an Material, die dabei entsteht, widersetzt sich den ‚klassischen‘ Mitteln der Literatur- und Kulturwissenschaft – vor allem dem ‚close reading‘. Vor ein vergleichbares Problem gestellt – nämlich angesichts der bis zu 60.000 Romane, die im England des 19. Jahrhunderts gedruckt worden sind, hat der Literaturwissenschaftler Franco Moretti vorgeschlagen, Texte ‚aus der Ferne‘ zu lesen. Statistiken wie etwa Worthäufigkeiten verknüpft mit mathematischen Visualisierungsstrategien werden im Rahmen des sogenannten ‚distant reading‘ dazu genutzt, grobe Zusammenhänge zwischen den Texten ‚herauszulesen‘, ohne sie im Detail studieren zu müssen. Franco Moretti plädiert dabei für eine Kombination von ‚distant reading‘ und ‚close reading‘. Mittels ‚Fernlesens‘ werden die Texte ausgewählt, mit denen man sich anschließend eingehender beschäftigen will, die man also ganz ‚klassisch‘ liest:

[W]e know how to read texts, now let's learn how *not* to read them. Distant reading: 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. And if, between the very small and the very large, the text itself disappears, [...] one can justifiably say, Less is more. (Moretti 2000:57, Hervorh. i. O.)

Allgemein sind quantitative Methoden mittlerweile in der Literatur- und Kulturwissenschaft angekommen. Texte stehen dabei eindeutig im Vordergrund, was nicht weiter verwundert, da Wörter zu zählen technisch sehr einfach zu bewerkstelligen ist. Doch auch für Bilder ist ein ähnlicher Zugang möglich (Kwastek 2013). Nachfolgend soll deshalb eine neue Methode vorgeschlagen werden, die es erlaubt, *YouTube*-Videos nach quantitativen Gesichtspunkten zu untersuchen: ‚distant watching‘ (‚Fern-Sehen‘). Als konkretes Anwendungsbeispiel soll das tschechoslowakische Musikvideo „Jožin z bažin“ („Seppl aus dem Sumpf“) aus dem Jahr 1978 dienen, das 2008 im Internet seinen Durchbruch gefeiert hat und zu einem Mem geworden ist.

2. Vorarbeiten

Entstanden ist die Idee zu ‚distant watching‘ im Rahmen eines Projekts zu Online-Erinnerungen an den umstrittenen ukrainischen Nationalhelden Stepan Bandera in Polen, Russland und der Ukraine (Fredheim/Howanitz/Makhortykh 2014). Dabei wurden zunächst die Aktivitäten von Nutzerinnen und Nutzern auf *Wikipedia*, *YouTube* und *Twitter* verglichen. Konkret wurden Zugriffszahlen, Geoinformationen, Artikel, Tweets und Kommentare quantitativ erfasst und verarbeitet. Um auch die Videoinformationen auf *YouTube* miteinbeziehen zu können, wurde von mir ein erster, rudimentärer Prototyp des ‚distant watching‘-Verfahrens entwickelt. Diesen Prototypen habe ich nun grundlegend überarbeitet und weiterentwickelt, um ihn auf eine große Bandbreite von Problemen der Internetkultur anwenden zu können.

3. Methodologie

Die Grundidee von ‚distant watching‘ ist, die ersten 200 Resultate für einen Suchbegriff – beispielsweise „Jožin z bažin“ – auf *YouTube* als Korpus zu verwenden. Die Einzelbilder (Frames) dieser 200 Videos dienen als statistische Grundlage. Zunächst werden die Unterschiede zwischen aufeinanderfolgenden Frames berechnet. Dabei entstehen charakteristische ‚Schnittkurven‘, die zeigen, dass sich alle Videos in fünf verschiedene (technische) Genres einreihen: geschnittene und ungeschnittene Videos, Videocollagen, Standbilder und Diashows. In den Schnittkurven fungiert die x-Achse als Zeitachse, die allerdings nicht Sekunden, sondern Einzelbilder erfasst. Auf der y-Achse werden die Unterschiede zwischen benachbarten Frames eingetragen, wobei 1,0 als 100% Übereinstimmung zu lesen ist, also kein Unterschied feststellbar ist, während 0,5 für 50% Übereinstimmung steht, etc. In Abbildung 1 sind Beispiele für alle fünf technischen Genres abgebildet. Sie stammen aus den *YouTube*-Suchresultaten für den Begriff „Jožin z bažin“ und zeigen, auf welche verschiedenen Weisen Nutzerinnen und Nutzer das Originalmaterial aufgreifen, bearbeiten und weiterverbreiten.

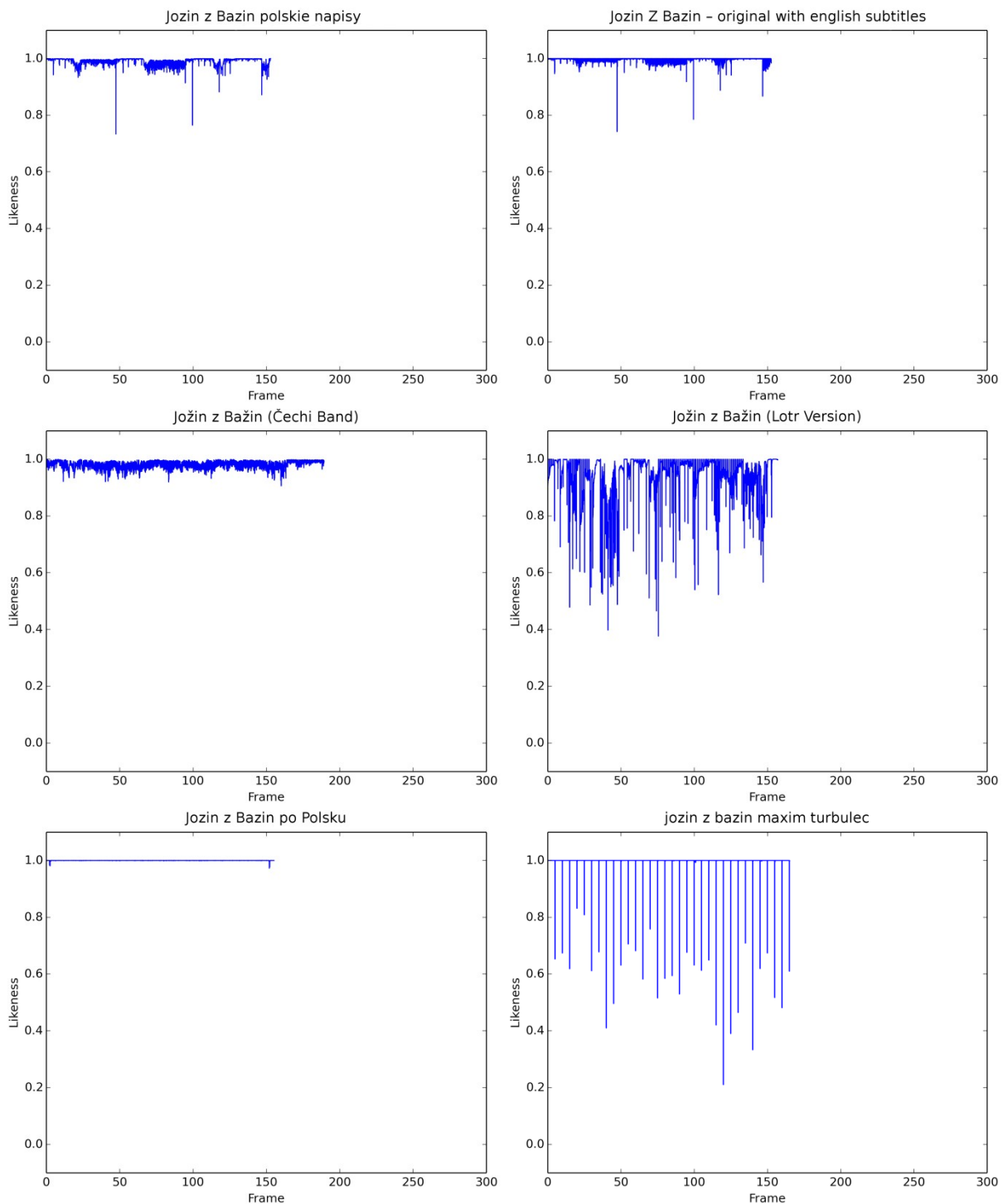


Abbildung 1. Schnittkurven zweier Versionen des (geschnittenen) Originalvideos (oben rechts und links), eines ungeschnittenen Amateurvideos (Mitte links), einer Collage aus unterschiedlichen Videos (Mitte rechts), eines Standbildes (unten links) und einer Diashow (unten rechts).

Die Schnittkurven beschleunigen die (manuelle) Analyse der 200 Videos im Korpus, unter Zuhilfenahme der Hauptkomponentenanalyse (*Principal Component Analysis*, PCA) kann man darüber hinaus einzelne Videos automatisiert zueinander in Bezug zu setzen. Die PCA erlaubt es, hochkomplexe Probleme kompakt in Form einer zweidimensionalen Grafik darzustellen. Dazu muss man allerdings die Daten in eine einheitliche Form bringen. Im Falle der Videos ist dies aufgrund ihrer unterschiedlichen Länge notwendig. Um die Videos zu vereinheitlichen, werden sie über die 70 größten Framesunterschiede in absteigender Reihenfolge codiert, wobei jeweils die ersten und letzten zehn Sekunden eines Videos nicht berücksichtigt werden. Diese Herangehensweise wurde experimentell ermittelt. Weiters muss berücksichtigt werden, dass der größte Framesunterschied in manchen Videos (Abb. 1 links unten) keine Bedeutung hat, weil *alle* Unterschiede 1,0 betragen. Deshalb wurde 0,9 als Schwellenwert eingeführt, ist der Framesunterschied kleiner, so wird stattdessen ‚-1‘ eingetragen. Damit sind die Daten hinreichend aufbereitet und man kann die PCA durchführen. Diese reduziert die ursprünglich 70 Dimensionen (die 70 größten Framesunterschiede) des Problems auf zwei, wodurch sich die 200 Videos ganz einfach darstellen lassen, siehe Abbildung 2.

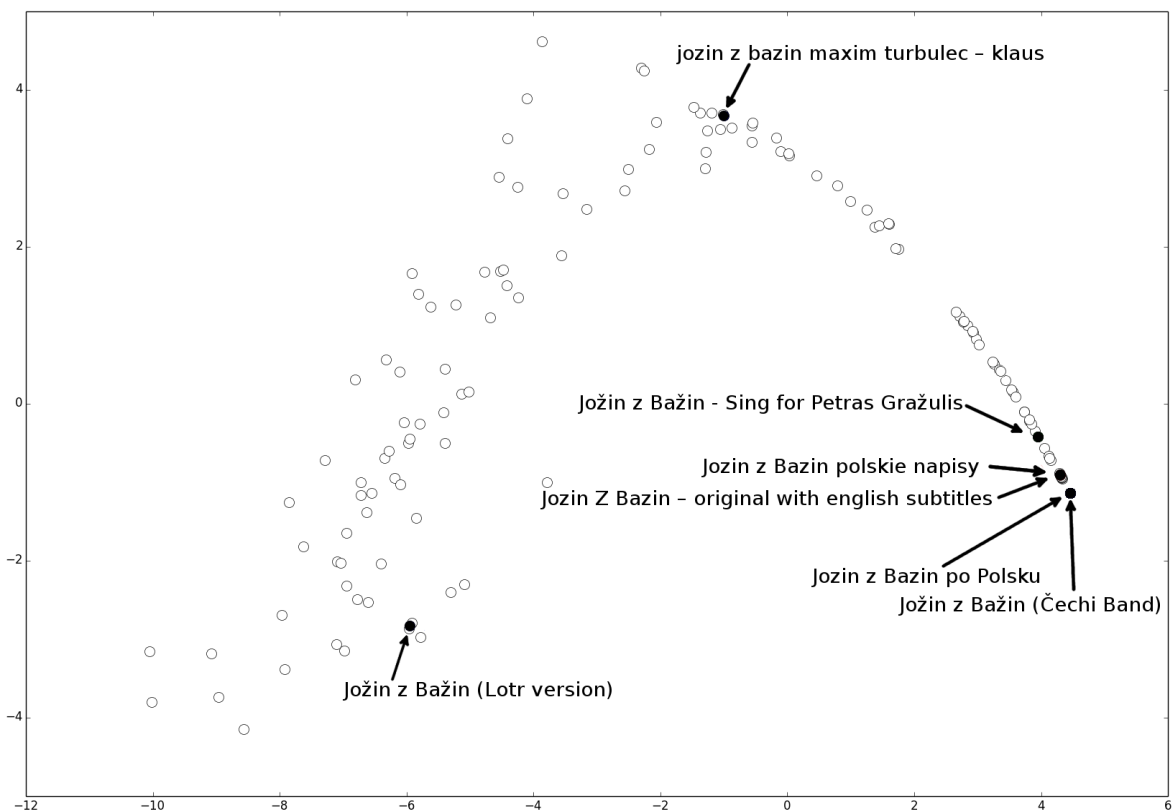


Abbildung 2. PCA der Einzelbilderunterschiede mit anschließender Projektion der Daten auf die von den ersten beiden Hauptkomponenten aufgespannte Ebene.

4. Ergebnisse

Die Schnittkurven sind ein gutes Hilfsmittel um das Genre eines einzelnen Videos auf einen Blick zu erfassen. So ist es offensichtlich, dass es sich bei den Videos oben links und rechts in Abbildung 1 um zwei unabhängig voneinander hochgeladene identische Versionen des originalen *Jožin z Bažin*-Videos handelt. Die Unterschiede ergeben sich aus der unterschiedlichen Videoqualität und den verwendeten Kompressionsverfahren. Die Schnittkurve in der Mitte links zeigt hingegen ein Video, das keine Schnitte enthält. Meist sind dies Amateurvideos. Rechts daneben sehen wir eine Videocollage, die passend zur *Jožin z Bažin*-Musik aus vielen sehr kurzen Videoschnipseln zusammengesetzt worden ist – hier konkret aus Peter Jacksons *Lord of the Ring: The Two Towers* (2002). Unten links ist ein Video, das nur aus einem einzelnen Standbild besteht, und unten rechts eine Diashow, die aus rhythmisch wechselnden Einzelbildern besteht. Da sich die technischen Genres hinsichtlich ihrer Aufwändigkeit unterscheiden, lassen sie erste Rückschlüsse bezüglich des Videoinhalts zu. Bei Standbildern ist es wahrscheinlich, dass die Musik im Vordergrund steht, weil die Videos optisch wenig zu bieten haben. Standbilder und Diashows sind einfach zu erstellen, geschnittene Videos hingegen anspruchsvoller. Dementsprechend selten sind selbst erstellte Videos unter den geschnittenen Videos; hier überwiegen Inhalte, die ursprünglich für das Fernsehen produziert worden sind.

Die Überblicksdarstellung der PCA erlaubt es, das Gesamtkorpus in den Blick zu nehmen. Wie man in Abbildung 2 sieht, fächert die PCA die Videos je nach Anzahl der Schnitte auf. Rechts unten finden sich sowohl das Amateurvideo der „Čechi Band“ als auch das Standbild *Jožin z Bažin po Polsku*. Nicht weit davon entfernt sind die zwei Versionen des Originalvideos – sie weisen relativ wenig Schnitte auf. Nach einer größeren Lücke in der ‚Perlenkette‘ kommt dann die Region der Videos mit häufigen Schnitten. Dementsprechend findet sich oben in der Mitte die Diashow *maxim turbulec*, und links unten die Videocollage *Lotr version*.

5. Diskussion

Der Masse an Material, welche die Netzkultur kreiert, kann man sinnvoll nur über quantitative Verfahren erfassen – beispielsweise mit ‚distant watching‘. Zwar gibt es hier noch Verbesserungspotential, etwa hinsichtlich der Erkennung ähnlicher Schnittfolgen, vor allem, weil es vielversprechend erscheint, diese Methode auf andere Probleme – etwa in der Filmanalyse – anzuwenden. Auch könnte eine quantitative Analyse der Tonspuren zusätzliche Einblicke bringen – etwa, ob das Originallied verwendet wird oder nicht. Statistische Informationen wie die Aufrufzahlen oder Anzahl der Kommentare auf *YouTube* können als zusätzliche Unterstützung verwendet werden. Auch können

die Frames mit Bilderkennungsalgorithmen weiter bearbeitet werden, um etwa das Auftreten bestimmter Symbolkombinationen quantitativ zu erfassen. All diese zusätzlichen Verfahren müssen natürlich auf die jeweilige konkrete Fragestellung abgestimmt werden.

Trotzdem liefert ‚distant watching‘ wesentliche Einblicke in das Videokorpus. Es legt die grundlegenden technischen Differenzen frei, aufgrund derer sich die fünf technischen Genres bestimmen lassen. Die graphische Darstellung der PCA kann dazu benutzt werden, sich einen Überblick über das Korpus in seiner Gesamtheit zu verschaffen und Videos eines bestimmten Genres schnell und unkompliziert zu finden. Sie zeigt Gruppen von Videos auf, die insofern interessant sind, als sie eigene thematische Akzente setzen. Zu betonen ist, dass erst eine qualitative Untersuchung einiger repräsentativer Videos bzw. Video-Gruppierungen genaueren Aufschluss über die innere Beschaffenheit des Korpus zulässt.

Quellen

- Dawkins, R. 2013. „Just for Hits – Richard Dawkins“, The Saatchi & Saatchi New Directors’ Showcase. <http://www.youtube.com/watch?v=GFn-ixX9edg> (aufgerufen am 10. 11. 2014).
- Fredheim, R. / Howanitz, G. / Makhortykh, M. 2014. „Stepan Bandera through the Lens of Quantitative Memory Studies“, in *Digital Icons* 12 [im Druck].
- Kwastek, K. 2013. „Vom Bild zum Bild: Digital Humanities jenseits des Texts“. Keynote DHd 2013. <http://video.uni-passau.de/video/Vom-Bild-zum-Bild%253A-Digital-Humanities-jenseits-des-Texts/92b59154b625a12070cc9d6947d6a611> (aufgerufen am 10. 11. 2014).
- Moretti, F. 2000. „Conjectures on World Literature“, in *New Left Review* 1, 54–68.
- Shifman, L. *Memes in Digital Culture*, Cambridge (MA) 2013.