

Empirische Aufmerksamkeitseffekte multimodaler Kohäsion im Film

Laubrock, Jochen

laubrock@uni-potsdam.de

Universität Potsdam; Medizinische Hochschule Brandenburg,
Germany

Tseng, Chiao-I

tseng@uni-bremen.de

Universität Bremen

Wir kombinieren die Diskursmethode multimodaler Kohäsion mit empirischen Daten zu Aufmerksamkeit und narrativem Verstehens. Multimodale Kohäsion bezieht systematisch die in auditiver, visueller und verbaler Modalität auftretenden Ereignisse auf modalitätsübergreifende Diskursstrukturen. Wir nutzen diese Diskursstrukturen, um daraus theoriegeleitet empirisch prüfbare Vorhersagen abzuleiten. Wir überprüfen mit Blickbewegungsexperimenten und Fragebogenstudien, wie kohäsive Hinweise Aufmerksamkeit und das Verständnis des Narrativs beeinflussen. Konkret haben wir mit Videobearbeitungssoftware kritische kohäsive Hinweisreize z.B. aus der Eingangsszene von Hitchcock's "The Birds" entfernt und mittels Eyetracking die Aufmerksamkeitsverteilung von insgesamt 114 Betrachtern gemessen. Unterschiedliche Gruppen von Probanden sahen Originale und manipulierte Versionen. Die kritischen kohäsiven Hinweisreize wurden im Original deutlich häufiger betrachtet als äquivalente Orte in der manipulierten Version. Also werden kohäsive Hinweise im Normalfall tatsächlich beachtet. Die Effekte kohäsiver Hinweise wirken nach: In einer anschließenden, für beide Versionen identischen Szene zeigten Betrachter ohne narrative Hinweise ein deutlich diffuseres Orientierungsverhalten. Narrative Elemente im Film lenken die Aufmerksamkeit des Betrachters.

Von verbaler zu multimedialer Kohäsion

Die Analyse von Kohäsionsmitteln ist seit langem etabliert als Methode der linguistischen Textanalyse (Halliday & Hasan 1976). Sie basiert auf der Beobachtung, dass z.B. Wiederholungen und Rekurrenzen linguistischer Muster funktional dafür sind, wie Sätze in einem Text als zusammenhängend betrachtet werden. Obwohl alle verbalen Texte verschiedene Formen von Kohäsionsmitteln nutzen, hat es sich bisher trotz einiger diesbezüglicher Corpusstudien (z.B. Flowerdew & Mahlberg 2009; Tanskanen 2006; Hoffmann 2012) als schwierig erwiesen, direkte Verbindungen zwischen verschiedenen Arten von Kohäsion und Diskursinterpretationen zu ziehen, beispielsweise bezüglich Anaphern oder Koreferenzketten.

Martin (1992: Chapter 3) entwickelte eine funktional organisierte Diskurssemantik für verbal Texte. Kohäsion wird betrachtet als Menge kommunikativer Ressourcen zur Präsentation und

Verfolgung von Diskursreferenzen über Text hinweg mit Fokus vor allem auf Menschen, Orten und Objekten und zur Klassifikation der Verbindungen korrespondierender Elemente durch spezifizierte Präsentations- und Verfolgungsstrategien. Dadurch dass Martin für Kohäsion explizit eine höhere Abstraktionsebene vorsieht, die sich von den spezifischen linguistischen Elementen und Formen unterscheidet, wird es leichter, semiotisch unterschiedliche (z.B. verbale und visuelle) Elemente in gemeinsame Diskursstrukturen zu integrieren.

Hier erweitern wir diese Linie auf das audiovisuelle Medium des Films. Film eignet sich besonders gut zur explorativen kohäsiven Analyse. Erstens kombinieren Filme nicht nur Text und Bild, sondern auch gesprochene und geschriebene Sprache, Klänge, Bewegungen und andere visuell getragene Information wie Betrachtungspunkte, Gestik, Mimik, Nähe etc. (Bordwell 2007) in einer bewusst integrativen Art und Weise. Zweitens sind Filme trotz ihrer semiotischen Komplexität immer noch primäre lineare expressive Formen, die sich normalerweise linear in der Zeit entfalten. Drittens gibt es zunehmend Arbeiten zu Gemeinsamkeiten der kognitiven und neuronalen Korrelate des Diskursverständnisses in Film und Text (Zacks & Magliano 2011; Zacks et al. 2007; Kurby & Zacks 2008; Radvansky & Zacks 2017). Aus experimentellen Studien wissen wir, dass Leser wie Filmbetrachter mit dem Ziel des Diskursverständnisses Orte, Zeiten, Handelnde und Kausalbeziehungen eng verfolgen. Wenn sich zentrale Merkmale der Situation ändern, wird dies als Ereignisgrenze wahrgenommen und das aktuelle Ereignismodell muss aufgefrischt werden (Zacks et al. 2009). Solche Merkmale gleichen denen, die in der Kohäsionsanalyse im Mittelpunkt stehen.

Multimodale Kohäsion als Bestandteil filmischen Diskurses

Filmwissenschaftler sind sich einig darüber, dass im audiovisuellen Medium des Films Wiederholungen, Rekurrenzen und Ähnlichkeiten in der Form systematisch eingesetzt werden, um es dem Betrachter zu erleichtern, zu kohärenten Interpretationen des Materials zu gelangen und ihn emotional und aesthetisch zu involvieren. Obwohl die allgemeine Konzeption von Kohäsion als Hinweis darauf, wie Diskursentitäten eingeführt und verfolgt werden können, beibehalten wird, existieren beim Film im Vergleich zum Text eine weitaus größere Zahl möglicher Kommunikationsmittel. Diese erweiterte Form der Kohäsion wird in Tseng (2013) im Detail diskutiert; wir nutzen sie hier zur Analyse spezifischer Filmsequenzen.

In diesem Rahmenmodell werden kohäsive Mechanismen als Strategien angesehen, den Betrachter zu bestimmten Interpretationen zu führen beim Versuch, Ereignisse in audiovisuellen Medien zu verstehen. Hier zeigen wir, wie Kohäsionsanalyse es uns ermöglicht zu verstehen, wie technische, beobachtbare Merkmale eines Filmes die Interpretation anleiten, während sich dynamische Ereignisse entwickeln. Die kohäsive Analyse beschreibt, wie Bild, Ton, verbale Sprache, geschriebene Sprache, Kamerabewegung, *Framing*, Farbe und viele weitere Aspekte es bewirken, dass Personen, Orte und Objekte in einer gegebenen Ereignissequenz eingeführt und verfolgt werden können.

Gibt es systematische empirische Korrelate der abstrakten Kohäsionsanalyse? Mit anderen Worten, stehen die tatsächlichen Prozesse des Diskursverständnisses in einem systematischen Zusammenhang mit den Mustern der kohäsiven Analyse? Hier zeigen wir, dass wir aus der abstrakten kohäsiven Analyse spezifischer Filmsequenzen Elemente ableiten können, die für das

weitere Verständnis des Narrativs von entscheidender Bedeutung sind. Wir untersuchen dies mit einem der Experimentalpsychologie entlehnten Ansatz der isolierten Bedingungsvariation. Wir manipulieren gezielt und subtil das Filmmaterial an Stellen, die gemäß der theoretischen Kohäsionsanalyse als besonders wichtig erscheinen, und beobachten die Effekte auf das Verständnis der Betrachter. Wir testen die Hypothesen mittels Fragebogendaten und Blickmessung (*Eye-Tracking*). Wir illustrieren dies am Beispiel einer kohäsiven Analyse der Anfangssequenz von Alfred Hitchcocks *The Birds* (Die Vögel, 1963), aus der wir theoriegeleitet Vorhersagen ableiten, die wir anschließend in Verhaltensexperimenten empirisch überprüfen.

Zu Beginn des Films wird die Hauptperson narrativ vom Hintergrund in den Vordergrund geholt. In der gesamten Außenszene hört man immer wieder kreischende Möwen. Der Film beginnt mit einer vorbeifahrenden Straßenbahn, die den Blick auf eine Gruppe von Menschen an einer belebten Kreuzung freigibt, welche darauf warten die Straße zu überqueren. Eine weibliche Person trennt sich von der Menge und die Kamera fokussiert auf sie, was sie visuell salient macht. Sie verschwindet hinter einem Poster der Golden Gate Bridge mit dem Schriftzug San Francisco und taucht dann wieder auf als Person im Vordergrund vor einer Tierhandlung. Man sieht, wie sie sich nach dem Geräusch der kreischenden Möwen am Himmel umdreht, woraufhin die Kamera auf den Möwenschwarm schwenkt. Nun wird wieder die Frau fokussiert, und man sieht, wie sie die Tierhandlung betritt (die zeitgleich Hitchcock verlässt). Schließlich wird sie in der Tierhandlung gezeigt, wie sie eine Treppe heraufsteigt.

Das Ergebnis der kohäsiven Analyse ist in Abbildung 1 dargestellt. Für die empirische Untersuchung ist besonders wichtig zu erkennen, dass schon frühzeitig, etwa in 6, der Hinweis auf die Tierhandlung als verbaler Hinweisreiz zu erkennen ist („Davidson's Pet Shop“). Dieser Hinweisreiz etabliert das Setting, nachdem die Handelnde durch die Eingangstür der Handlung gegangen ist und erleichtert es dem Betrachter, Kohäsion wahrzunehmen.

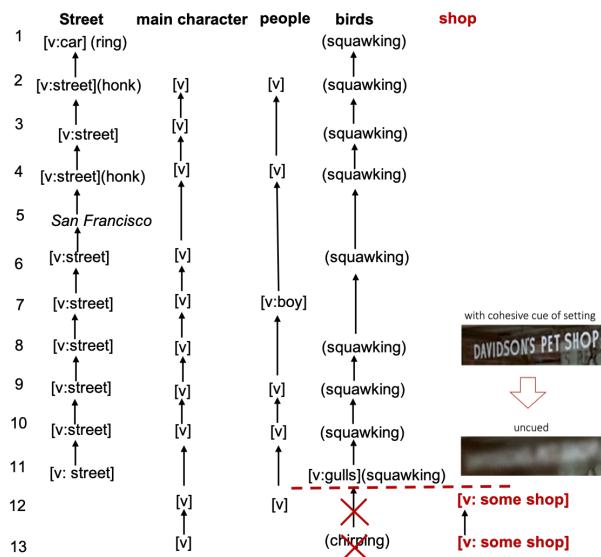


Abb. 1: Kohäsionsanalyse (schwarz) und Veränderungen in der manipulierten Version (rot) der Eingangssequenz von Hitchcock's *The Birds*.

Empirische Überprüfung

Wenn wir nun Hinweisreize auf die Tierhandlung entfernen, sollte dem Betrachter die Interpretation des *Settings* schwerer fallen. Wir haben den Film subtil verändert, indem wir die Schilder durch Tiefpassfilterung unleserlich gemacht haben. Zusätzlich haben wir die Vogelgeräusche im Inneren der Tierhandlung durch generische sanfte Hintergrundmusik ersetzt. Die Sequenz innerhalb der Tierhandlung war somit visuell identisch für das Original und die manipulierte Version. Obwohl die Manipulation flüchtigen Betrachtern nicht auffiel, erwarteten wir durch die Zerstörung kohäsiver Ketten einen Effekt auf das narrative Verständnis der Szene innerhalb der Tierhandlung. Die roten Elemente in Abbildung 1 zeigen, dass das Entfernen der visuell-verbalen Hinweise auf die Tierhandlung das Setting von einer Tierhandlung zu einem generischen Geschäft änderte. In ähnlicher Weise sollte sich das Ersetzen des Vogelzwitschers durch Aufzugsmusik auswirken. Dadurch dass der Betrachter weniger vorbereitet ist, erwarteten wir mehr Orientierungsverhalten.

In einer Fragebogenstudie erfragten wir unmittelbar im Anschluss an das Filmbetrachten das offene Wissen um das Setting: „Wohinein ging die handelnde Person?“. In zwei unterschiedlichen Stichproben in Bremen (n=45) und in Potsdam (n=74) gab es deutliche Unterschiede zwischen den Gruppen, die das Original und die manipulierte Version gesehen hatten ($p = 0.009$ bzw. $p = 0.033$). Jeweils war die Anzahl richtiger Antworten höher in der Originalversion.

In einer Blickbewegungsstudie haben wir zunächst geschaut, ob sich das Entfernen der kohäsiven Hinweisreize auf die Aufmerksamkeitsverteilung ausgewirkt hat. Abbildung 2 zeigt, dass dies eindeutig der Fall war. Wenn Schrift zu lesen war, war ein nennenswerter Anteil (18%) aller Fixationen auf die Schilder fokussiert, die also stark aufmerksamkeitslenkend wirkten. Wenn sie dagegen verschwommen dargestellt wurden, wurden sie kaum beachtet (1% der Fixationen). Dieser Gruppenunterschied war hoch signifikant, $p < 0.001$. (3% der Probanden mit Hinweisreiz, aber nur 15% der Probanden ohne Hinweisreiz fixierten die Schilder mindestens einmal.

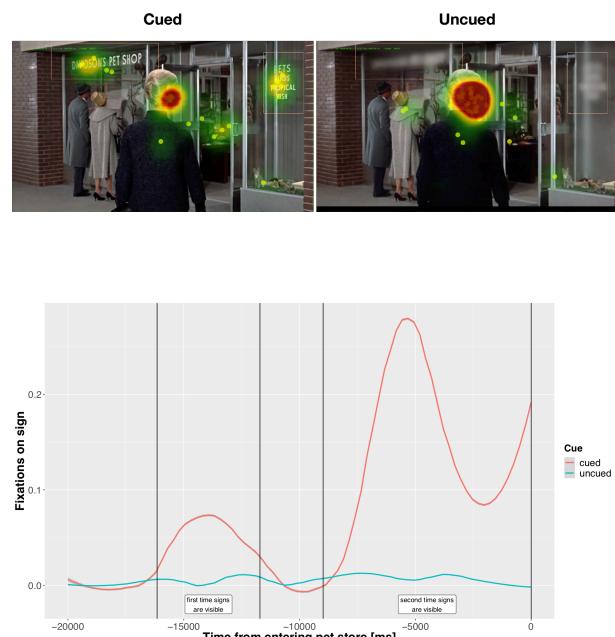


Abb. 2: Anzahl der Fixationen auf den Schildern in der Version mit und ohne Hinweisreiz (cued vs. uncued), relativ zum Eintreten in die Tierhandlung (Zeit = 0).

In einem weiteren Schritt haben wir untersucht, ob das Fehlen kohäsiver Hinweise zur Etablierung des Tierhandlungs- *Settings* zu mehr Orientierungsverhalten bei der visuell identischen Szene *innerhalb* der Tierhandlung führten. Dazu haben wir ein Maß benutzt, das als Erweiterung der Standardabweichung auf zwei Dimensionen angesehen werden kann (die Quadratwurzel der Determinante der Kovarianzmatrix, Paindaveine 2008). Abbildung 3 zeigt wie sich die Streuung der Blickverteilung über die Zeit in der Tierhandlung entwickelt. In beiden Bedingungen steigt die Streuung anfangs und sinkt gegen Ende der Szene. Jedoch beginnt das Absinken in der Bedingung mit Hinweisreiz früher und es ist deutlich. Der Beginn des Absinkens scheint mit dem Moment übereinzustimmen, an dem die Handelnde ihren Kopf dreht. Wir spekulieren, dass das Kopfdrehen ein soziales Signal für den Betrachter ist, ihrer Aufmerksamkeit zu folgen, und dass dieser Cue in der Originalversion häufiger beachtet wurde, während Betrachter der manipulierten Version weiter das Geschäft explorierten. Über die gesamte Szene aufsummiert ist die Streuung der Verteilung eindeutig größer in der Bedingung ohne Hinweisreiz, obwohl sich die Verteilungen zu Beginn mehr oder weniger parallel entwickeln.

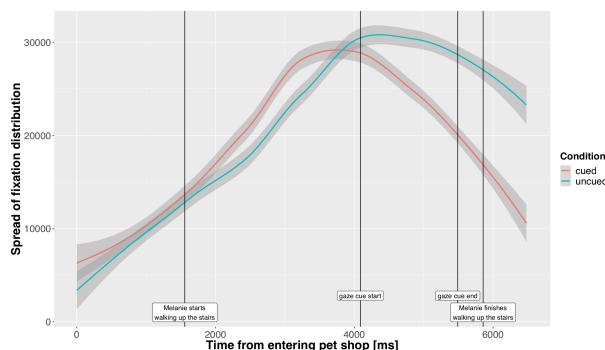


Abb. 3: Entwicklung der Streuung der Blickverteilung innerhalb des Geschäfts über die Zeit.

Während die Fragebogendaten zeigen, dass die Teilnehmer weniger wahrscheinlich den Ort identifizieren, wenn kohäsive Hinweise fehlen, legen die Blickbewegungsdaten nahe, dass Betrachter des Originals tatsächlich die "Pet Shop"-Schilder lesen und dazu nutzen, die Identität des Geschäfts zu etablieren, während Betrachter der manipulierten Version später innerhalb des Geschäfts aktiv nach Information suchten, um ihre Unsicherheit zu reduzieren. Zusammenfassend betrachtet haben hier verbale und auditive Hinweise die Aufmerksamkeit des Betrachters gelenkt und dadurch ihren narrativen Verstehensprozess beeinflusst.

Diskussion

Die Kohäsionseffekte waren relativ klein gemessen an dem Effekt der visuellen Manipulation. Dies hatten wir erwartet vor dem Hintergrund dass Aufmerksamkeit sowohl durch den Reiz als auch durch höhere Ziele gelenkt werden kann, und die im Film dominanten Bewegungssignale sehr stark reizgetriebene die Aufmerksamkeit aus sich ziehen. Tatsächlich ist diese reizgetriebene, bottom-up Steuerung der Aufmerksamkeit und des Blickes bei

gleichzeitig nur schwer detektierbaren top-down-Effekten so ausgeprägt, dass sie als "tyranny of film" bezeichnet wurde (Loschky et al. 2015).

Auch in unserer Studie drückten sich solche reizgetriebenen Effekte aus in der Homogenität der Blickverteilung und vor allem auch im starken Effekt, den unsere Manipulation hinsichtlich der Betrachtung der manipulierten Reize selbst hervorrief (vgl. Abb. 2). Nichtsdestotrotz haben wir auch kleinere, zielgesteuerte top-down-Effekte auf die Aufmerksamkeitsverteilung nachweisen können, die durch Unterschiede in der Kohäsion verursachte Unterschiede im Verstehen der Szenesemantik widerspiegeln. Letztlich führten diese zu Unterschieden der Aufmerksamkeitsverteilung auf visuell identischen nachfolgenden Szenen.

Inwieweit liefert dieser Beitrag einen Mehrwert für die DH? Wir haben digitale Methoden verwendet bei der Bearbeitung der Film-ausschnitte, der Messung der Blickdaten und bei der Erstellung von R-Skripten für Auswertungen und Visualisierungen. Jedoch haben wir hier keinen *Distant Viewing*-Ansatz mit rein informatischer Inhaltsanalyse, sondern eher einen *Close Viewing*-Ansatz verfolgt (im Sinne des *Close Reading*). Rein computerbasierte Inhaltsanalysen mit tiefen neuronalen Netzen, wie sie von uns etwa bei der stilometrischen Analyse graphischer Romane implementiert wurde, liefern sicher perspektivisch wichtige zusätzliche Erkenntnisse; sie sind aber derzeit noch nicht auf dem Niveau entwickelt, mit dem sich hypothesegeleitete semantisch-narrative Fragestellungen wie die unsrige mit vertretbarem Aufwand sinnvoll verfolgen ließen. Unsere Studie werten wir eher als eine erste explorative empirische Vorarbeit, die Fragestellungen für spätere, möglicherweise „digitalere“ Analysen aufzeigen kann.

In diesem Beitrag haben wir einen theoretischen Rahmen multimodaler Diskursanalyse verknüpft mit einer empirischen Überprüfung von daraus abgeleiteten Vorhersagen. Wir haben gezeigt, dass kohäsive Hinweise zum Verständnis des Films beitragen und dem Betrachter bei der Interpretation helfen. Wir wissen nun etwas besser, wie kohäsive Hinweise im Film funktionieren. Spannend für zukünftige Forschung wäre die Frage, ob sich daraus auch Konsequenzen für Filmschaffende ableiten lassen.

Bibliographie

- Bordwell, D.**, 2007, *Poetics of Cinema*, Routledge, London and New York.
- Flowerdew, J.** and **Mahlberg, M.**, eds (2009), *Lexical Cohesion and Corpus Linguistic*, John Benjamins, Amsterdam.
- Halliday, M. A. K.** and **Hasan, R.**, 1976, *Cohesion in English*, Longman, London.
- Hoffmann, C. R.**, 2012, *Cohesive Profiling: Meaning and Interaction in Personal Weblogs*, John Benjamins, Amsterdam.
- Kurby, C. A.** and **Zacks, J. M.**, 2008, ‘Segmentation in the perception and memory of events’, in: *Trends in Cognitive Sciences* 12, 72–79.
- Loschky, L. C.**, **Larson, A. M.**, **Magliano, J. P.** and **Smith, T. J.**, 2015, ‘What would jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension’, in: *PLOS ONE* 10(11), 1–23.
- Martin, J. R.**, 1992, *English text: Systems and structure*, Benjamins, Amsterdam.
- Paindaveine, D.**, 2008, ‘A canonical definition of shape’, in: *Statistics & Probability Letters* 78(14), 2240–2247.
- Radvansky, G. A.** and **Zacks, J. M.**, 2017, ‘Event boundaries in memory and cognition’, in: *Current Opinion in Behavioral Sciences* 17, 133–140.

Tanskanen, S.-K., 2006, Collaborating towards Coherence: Lexical Cohesion in English Discourse, John Benjamins, Amsterdam.

Tseng, C., 2013, *Cohesion in Film: Tracking Film Elements*, Palgrave Macmillan, Basingstoke.

Zacks, J. M., Speer, N., Swallow, K., Braver, T. and Reynolds, J., 2007, 'Event perception: a mind/brain perspective', in: *Psychological Bulletin* 133, 273–293.

Zacks, J. M. and Magliano, J. P., 2011, "Film, Narrative and Cognitive Neuroscience", in: F. Bacci and D. P. Melcher, eds, *Art and the Senses*, Oxford University Press, Oxford and New York, pp. 435–454.