

## „Computational literary studies“, 10 July 2014, Digital Humanities Conference 2014

### **Paper Session 4**

Chair: Jan Rybicki

#### **Beyond style: literary capitalism and the publishing industry**

James O'Sullivan (Cork, PA, USA) and Simon Fuller (Maynooth, Ireland)

Link zum Paper: <http://dharchive.org/paper/DH2014/Paper-179.xml>

#### **Progress through Regression. Modeling Style across Genre in French Classical Theater**

Christof Schöch (Würzburg, Germany) and Allen Ridell (Dartmouth, NH, USA)

Link zum Paper: <http://dharchive.org/paper/DH2014/Paper-60.xml>

#### **Making Waves: Algorithmic Criticism Revisited**

David L. Hoover (New York, NY, USA)

Link zum Paper: <http://dharchive.org/paper/DH2014/Paper-766.xml>

### **Report: Sebastien Dupont**

Within the field of Literature, there is a growing interest in applying computational techniques. At the same time, a subfield of Computational Linguistics is developing, that addresses a range of problems in the field of Literature. This session's contributors demonstrate the range of possible usages of computational methods, at the meeting points of those two fields of study. In this report, we will see such techniques in action in three case studies. The first one aims to identify the proportion of actual writing that is done by each co-author in a collective literary work. The second one traces the influence of the genre on author writing style in french classical theatre. The third one analyses the different character voices in Virginia Woolf's "The Waves" monologues.

### **Beyond style: literary capitalism and the publishing industry**

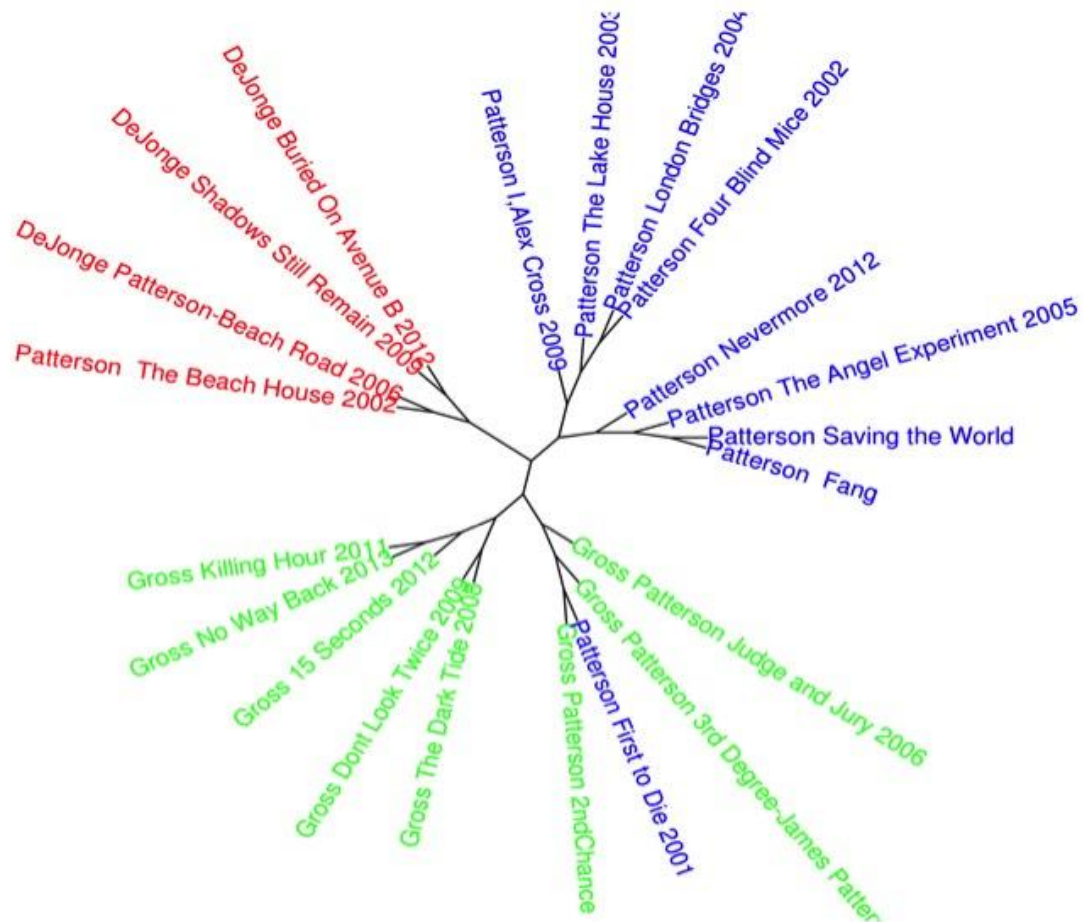
Pioneer of book advertising on television, the prolific and successful author [James Patterson](#) approaches the literary production process in a collaborative way: "I'll write an elaborate outline, maybe 70 pages, very detailed, clear, and focused. The co-author will write the first draft, and I'll see the work every few weeks. I'll do two to seven more drafts."<sup>1</sup>

**JAMES O'SULLIVAN** (Cork, PA, USA) and **SIMON FULLER** (Maynooth, Ireland) are looking at how much of the actual writing is done by every co-author. Using a "bootstrap consensus tree" cluster analysis over maximum frequency words with the Burrow's Delta metric, the novels can be classified by main authors (see Fig. 1). The data suggest that the bulk of the actual writing is done by Patterson's collaborators, which is often observed in "literary assembly lanes"<sup>2</sup>. We can trace back the industrialization of the literary production process to famous names such as Alexandre Dumas. Patterson provides a striking example where the capitalist system has been applied to produce [books](#) with plot-oriented storytelling in a very efficient way.

---

<sup>1</sup> **Patterson, James.** (2012) *Life's Work: James Patterson*. Harvard Business Review. Web.

<sup>2</sup> **Deighton, John.** (2006) *Marketing James Patterson*. Case Study. Boston. Harvard Business Publishing. Web. 24 August 2013. pp 4.



100-1000 MFW Culled @ 0%  
Classic Delta distance Consensus 0.5

Fig. 1 - Classification of novels by main author

Source: <http://dharchive.org/paper/DH2014/Paper-179.xml>

## Progress through Regression. Modeling Style across Genre in French Classical Theater

In French classical theater, the genre (e.g. comedy, tragedy, tragi-comedy, ...) is governed by a set of formal and stylistic [constraints](#). **CHRISTOF SCHÖCH** (Würzburg, Germany) and **ALLEN RIDELL** (Dartmouth, NH, USA) analyse the influence of the genre on the style of the author in french classical theatre plays<sup>3</sup>. They compare three models that predict the author of a section in a play: first based on word frequency alone, then adding the likelihood of an author writing in each genre and finally based on word frequencies for each genre separately (see Fig. 2). This analysis suggests that the genre can be a useful factor to perform authorship attribution, and also shows the importance of logistic regression in modelling a range of information jointly with authorship.

<sup>3</sup> Fièvre, P., ed. (2007-2013). *Théâtre classique*, [www.theatre-classique.fr/](http://www.theatre-classique.fr/).

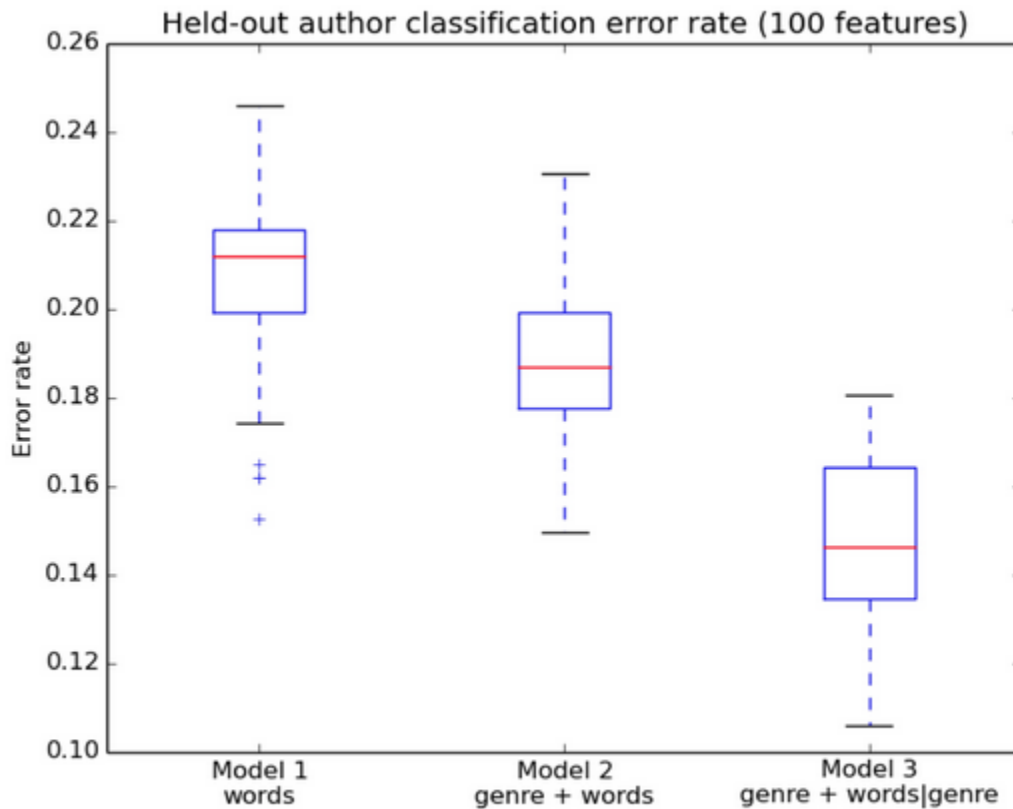


Fig. 2: Held-out author classification error rate (100 features)

Source: <http://dharchive.org/paper/DH2014/Paper-60.xml>

### Making Waves: Algorithmic Criticism Revisited

Virginia Woolf's probably most experimental book "The Waves" consists of soliloquies spoken by the book's six characters. **DAVID L. HOOVER** (New York, NY, USA) is trying to recreate Ramsay's analysis of character individualization of those six distinct character voices. Various author determination techniques<sup>4</sup>, such as tf-idf (See Fig. 3), word frequency list, 2-grams or Zeta<sup>5</sup> confirm that there is a stylistic voice for each of Virginia Woolf's characters. The usage of algorithms to analyse literary works also raises a very interesting problematic for the literary critics: identifying [computationally-tractable](#) and computationally-intractable questions and the significance of the boundary between those.

<sup>4</sup> Craig, H., and A. Kinney. (2010). Shakespeare, Computers, and the Mystery of Authorship. Cambridge UP.

<sup>5</sup> This method measures consistency of use rather than frequency.

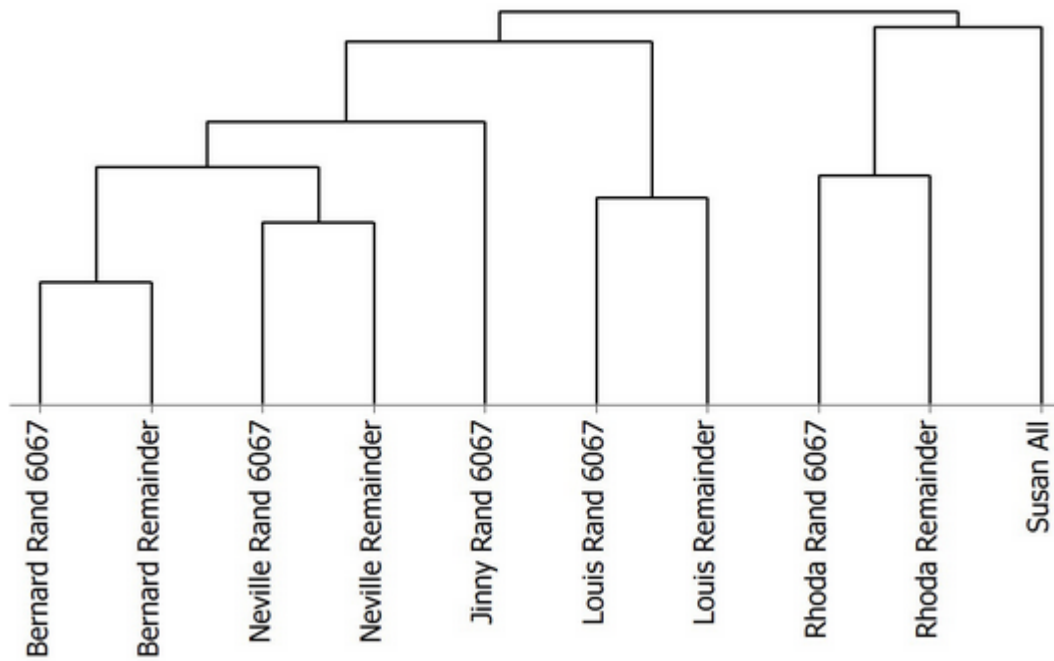


Fig. 3 - Tf-idf and Character Individualization  
Source: <http://dharchive.org/paper/DH2014/Paper-766.xml>

Those interesting applications of authorship determination techniques clearly show that some of the questions that the literary scholar may encounter can be properly answered with the help of computational techniques. Since questions have the tendency to keep popping up, those tools will continue to prove very valuable in the future.

Sebastien Dupont  
dupont.sebastien@gmail.com