Quality assessment of literature is a highly contested matter. Positions in the debate range from constructivist context dependency (‘the success of a work of literature depends entirely on its context’) to work internalism (‘success depends on work-internal features’), [1, 2]. While context dependency is evidenced by the variety and seemingly chaotic dynamics of bestseller lists, the constructivist argument ignores the convergence of an empirical ‘canon’ for many readers over time and space [3, 4]. This talk presents evidence from computational narratology that supports the claim that perception and transmission of literature tend to facilitate the emergence of specific properties of successful literature. We extract narrative arcs using rule-based sentiment analysis from 9000 novels annotated for success. The arcs’ global and local dynamic properties are subsequently estimated with a combination of multifractal and information-theoretic approaches, see Fig. 1. We find that the narrative structure of successful literature displays a tension between local uncertainty (i.e., segments of successful arcs show higher approximate entropy) and global coherence (i.e., successful arcs display long-range dependencies). We speculate that this tension between the local and global organization of language reflects that literature, like other cultural artifacts, has to balance attention and motivation in order to be successfully culturally transmitted [5, 6].

Figure 1: The narrative arc for noble laurate Kazuo Ishiguro’s Never Let Me Go - an example of a successful novel [5]. The upper row, left, in blue, is the raw sentiment score, and green and red are progressively smoothed using an adaptive filter at window sizes 55 and 735 sentences (Novel sentence length $L$ divided by 200 and 15). Right, the arc is scaled to $\{-1, 1\}$ to enhance the story arcs. Raw sentiment times series appear highly irregular, while adaptive filtering enhances abstract and recognizable story arcs. The story arc is particularly recognizable in the middle row left and right where black curves are smoothed at $w = 2763$ to extract the narrative’s abstract “tragedy” representation (i.e., negative beginning, positive middle, negative ending). The lower row shows that the arc, original left and rescaled right, display long-range dependencies at short and intermediary time-scales as evidenced by the Hurst exponent, $H_s$, which is an ‘index of long-range dependence’. It also shows that the arc is multifractal where the Hurst exponent varies with the time scale, ex. $H_s$ and $H_l$. 
References


