Big Data, Short Works: Establishing a Stylometric Baseline for Micro-attributions of Shakespeare’s Apocrypha with ‘On a day, alack the day’

Chance Swaim*
Faculty: Francis X. Connor
Department of English, Fairmount College of Liberal Arts and Sciences

With the digitization of the Early Modern dramatic canon nearly complete, the contemporary understanding of authorship is rapidly changing. Utilizing digital databases such as LION (Literature Online), which makes primary-source texts searchable, authorship can be determined more assertively by studying stylometric variations and matches by comparing large texts by authors with comparable corpus sizes. Big data allowed the editors of The New Oxford Shakespeare (2016) to attribute sections of plays traditionally thought to be authored by Shakespeare to other playwrights of the period, like Christopher Marlowe. But shorter, more unstable texts—such as Early Modern poetry, which is often relatively short, formal and circulated in manuscript form—present a more complex problem. My research looked at rare and unique phrases within the poem “On a day—alack the day!” (Love’s Labour’s Lost and The Passionate Pilgrim 1598) to establish a baseline statistical model for Shakespeare’s authorship of poems of comparable length.