

Diplomatic Vocabulary in Neville and Shakespeare

A systematic prevalence study of lemma frequency, work-level dispersion, chronology, and contemporary drama controls

Central result. A twenty-one-term diplomatic panel selected without using Shakespeare counts or dates occurs at 2.64 times the comparison-drama rate in Shakespeare's 1601–1603 works. Its rate in those works is 2.59 times Shakespeare's own pre-1599 rate. The effect belongs to the coordinated register and its timing, not to the rarity of any single word.

31 July 2026

Standalone research report

Contents

1	Research question and answer	1
2	Corpora and exact era definitions	1
2.1	Era rules	1
2.2	Corpus totals by era	1
2.3	Public source identities	2
3	Text preparation	2
4	Statistical methods	2
4.1	Frequency and dispersion	2
4.2	Work-level uncertainty	2
4.3	The twenty-one-term panel	3
4.4	Bootstrap and matched-panel calibration	3
5	Results	3
5.1	The panel's chronological profile	3
5.2	Aggregate contrasts	5
5.3	Individual panel terms	5
5.4	What the individual-word screen shows	6
5.5	Matched-panel calibration	6
6	Interpretation	6
7	Limitations	7
8	Conclusion	7
A	Exact Shakespeare works counted in each era	8
B	Exact comparison-corpus plays counted in each era	10
C	Reproducibility specification	17

1 Research question and answer

This study asks whether words concentrated in Sir Henry Neville’s diplomatic correspondence are unusually prevalent in Shakespeare, whether their prevalence changes around 1599–1603, and whether the same behavior occurs in other drama of the period. The statistical object is not an isolated lexical coincidence. It is the frequency, dispersion, chronology, and coordinated movement of a diplomatic-administrative vocabulary. The answer is qualified but clear. Individual terms are often too sparse or too polysemous to carry the claim by themselves. A pre-specified twenty-one-term panel, however, rises sharply in Shakespeare after the pre-1599 baseline, peaks in 1601–1603, and occurs there at substantially more than twice the rate found in contemporary comparison drama.

2 Corpora and exact era definitions

2.1 Era rules

Each work is assigned to exactly one era from its composition year in the analytical metadata. The dates are conventional scholarly estimates and should not be read as exact days or publication dates.

Table 1: Era boundaries used throughout the study

Era label	Included years	Interpretive role
Before 1599	1590–1598	Shakespeare baseline before Neville’s surviving embassy-era concentration
1599–1600	1599–1600	Immediate transition and Neville’s French embassy
1601–1603	1601–1603	Peak Shakespeare panel rate
1604–1615	1604–1615	Later comparison and persistence check

2.2 Corpus totals by era

The Shakespeare corpus contains plays, collaborative plays, poems, and the sonnet sequence. These non-dramatic works are retained because the question concerns the Shakespeare canon as a whole. The exact membership and work type of all forty-two works appear in Appendix A. The comparison corpus contains plays only; all 203 are listed in Appendix B. The Shakespeare token stream contains 18 plays

Table 2: Exact work and token denominators

Era	Shakespeare		Comparison drama	
	Works	Tokens	Plays	Tokens
Before 1599	20	418,093	39	634,764
1599-1600	3	67,455	19	393,359
1601-1603	6	121,110	27	472,918
1604-1615	13	294,102	118	2,224,464
Total	42	900,760	203	3,725,505

and two poems before 1599; three plays in 1599–1600; five plays and one poem in 1601–1603; and eight plays, four collaborative plays, and the sonnet sequence in 1604–1615. Full token streams, including stage

directions in plays, contribute to the denominators and lemma counts. The comparison set is operationally defined by excluding database records matched to First Folio plays. It is not a pure authorship category: *Pericles* and *The Two Noble Kinsmen* occur in the comparison set and also in the Shakespeare set as collaborative plays. Accordingly, this report uses “comparison drama,” not “non-Shakespeare drama,” as the reader-facing label. Appendix B makes the complete membership, including this overlap, explicit.

2.3 Public source identities

Shakespeare work texts and public access are documented by the Folger Shakespeare Library’s [downloadable works](#). The comparison-drama frame is documented by the Folger [Early Modern English Drama](#) project. The wider print context is documented by [EarlyPrint](#) and the [Text Creation Partnership](#). Neville’s public documentary frame is the National Archives’ [State Papers Foreign](#) guide and the public scan of Winwood’s *Memorials of Affairs of State*.

3 Text preparation

The Neville corpus contains 164 letters and three supplementary documents, totaling 135,770 word tokens. The Shakespeare corpus contains 900,760 tokens, and the comparison corpus contains 3,725,505 tokens. Every token is compared by lower-case lemma rather than surface spelling. This choice prevents editorial modernization in Winwood-derived texts from being mistaken for Neville’s orthography. Spelling-specific inference is not part of this prevalence study. The tested lexical inventory retains alphabetic lemmas of at least three characters and removes a declared function-word list. All tokens remain in the corpus denominators. This produces 16,850 candidate lexical lemmas. Proper names and foreign words remain visible in the unrestricted exploratory output, so interpretation relies on the pre-specified diplomatic set rather than an unreviewed top-word list.

4 Statistical methods

4.1 Frequency and dispersion

For lemma j in corpus g , prevalence is reported as $10^6 C_{gj}/T_g$, where C_{gj} is the total lemma count and T_g is the corpus token count. A rate ratio divides the two corpus rates. Dispersion records the number of whole letters, documents, or works containing the lemma.

4.2 Work-level uncertainty

Whole texts are the independent units. For each corpus, the expected count in work i is the corpus rate multiplied by that work’s token length. A sandwich variance uses the squared work-level residuals rather than treating every token as an independent observation. Formal word tests require at least two tokens distributed across at least two units in both corpora. Zero-cell ratios receive a half-token descriptive correction but no formal test. Two-sided word-level tests are corrected within each comparison family by the Benjamini–Hochberg false-discovery-rate procedure [1]. The full 16,850-lemma screen is exploratory. The coordinated panel analysis is the primary hypothesis test.

4.3 The twenty-one-term panel

Panel selection was completed without consulting Shakespeare counts, Shakespeare dates, or Shakespeare work identities. A lemma had to: (1) belong to a declared diplomatic/state semantic field; (2) occur at least five times in Neville; (3) occur in at least five Neville targets; (4) first survive in Neville in 1599 or 1600; and (5) be at least fifteen times as concentrated in Neville as in comparison drama. These rules yield:

allowance, advertise, advertisement, audience, avow, commission, commissioner, communication, difficulty, insist, mediation, memorial, negotiation, omission, overture, parliament, passport, proposition, propose, secretary, treaty.

Because comparison drama contributes to panel selection and later serves as a baseline, the panel test is internally controlled rather than a fully independent external replication. Shakespeare remains blinded during selection.

4.4 Bootstrap and matched-panel calibration

The principal panel intervals are checked with 20,000 bootstrap replicates resampling whole works, with seed 1599. A separate post hoc calibration constructs 20,000 random twenty-one-term panels matched on Neville frequency, Neville dispersion, first Neville year, comparison-drama frequency, and Neville/comparison enrichment. Shakespeare information is excluded from matching. This calibration tests whether the panel’s behavior is generic among Neville-concentrated words; it is a sensitivity analysis, not a preregistered confirmation.

5 Results

5.1 The panel’s chronological profile

Figure 1 displays the decisive relationship. Shakespeare and comparison drama begin at broadly similar panel rates. Shakespeare rises in 1599–1600, reaches 396.3 occurrences per million tokens in 1601–1603, and then recedes. Comparison drama stays near 125–155 per million throughout.

Table 3: Panel counts and rates by corpus and era

Corpus	Era	Works	Tokens	Panel	Rate/1m
Comparison drama	Before 1599	39	634,764	82	129.2
Shakespeare	Before 1599	20	418,093	64	153.1
Comparison drama	1599-1600	19	393,359	49	124.6
Shakespeare	1599-1600	3	67,455	14	207.5
Comparison drama	1601-1603	27	472,918	71	150.1
Shakespeare	1601-1603	6	121,110	48	396.3
Comparison drama	1604-1615	118	2,224,464	344	154.6
Shakespeare	1604-1615	13	294,102	77	261.8

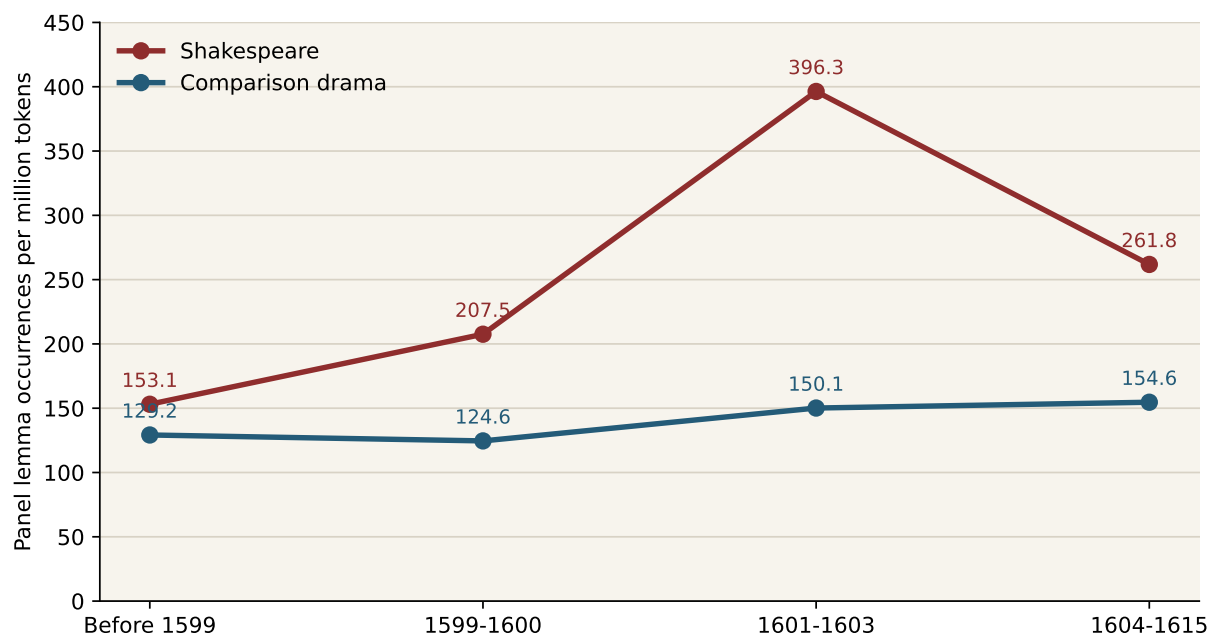


Figure 1: Twenty-one-term diplomatic panel by era. Values are occurrences per million lemma tokens. Exact works and token denominators are listed in Appendices A and B.

Table 4: Primary panel comparisons with work-robust uncertainty

Comparison	Rate ratio	95% interval	<i>p</i>
Shakespeare vs comparison drama, all years	1.54	1.12–2.11	0.0080
Shakespeare 1599–1603 vs Shakespeare before 1599	2.15	1.29–3.58	0.0033
Shakespeare 1601–1603 vs Shakespeare before 1599	2.59	1.59–4.22	0.0001
Shakespeare vs comparison drama, before 1599	1.18	0.68–2.07	0.5498
Shakespeare vs comparison drama, 1599–1600	1.67	1.17–2.38	0.0048
Shakespeare vs comparison drama, 1601–1603	2.64	1.75–3.99	< 0.00001
Shakespeare vs comparison drama, 1604–1615	1.69	0.94–3.06	0.0819

5.2 Aggregate contrasts

The whole-work bootstrap gives a 95% interval of 1.72–4.07 for Shakespeare versus comparison drama in 1601–1603 and 1.60–4.50 for Shakespeare 1601–1603 versus its pre-1599 baseline. The convergence between robust and bootstrap estimates makes the result less dependent on one uncertainty calculation.

5.3 Individual panel terms

The table below explains why the result must be understood as a coordinated register. Most terms are too sparse to support individual inference. Even where Shakespeare’s descriptive rate exceeds comparison drama, the uncertainty is often wide.

communication N 10 in 8 targets; S 2 in 2 works; C 0; S/C rate ratio 16.54; panel $q = \text{—}$.

treaty N 202 in 58 targets; S 7 in 4 works; C 16; S/C rate ratio 1.81; panel $q = 0.641$.

omission N 5 in 5 targets; S 1 in 1 works; C 0; S/C rate ratio 8.27; panel $q = \text{—}$.

advertisement N 70 in 40 targets; S 4 in 3 works; C 7; S/C rate ratio 2.36; panel $q = 0.539$.

advertise N 92 in 57 targets; S 10 in 7 works; C 15; S/C rate ratio 2.76; panel $q = 0.176$.

commissioner N 40 in 30 targets; S 1 in 1 works; C 12; S/C rate ratio 0.34; panel $q = \text{—}$.

difficulty N 42 in 29 targets; S 5 in 4 works; C 13; S/C rate ratio 1.59; panel $q = 0.641$.

negotiation N 13 in 12 targets; S 1 in 1 works; C 4; S/C rate ratio 1.03; panel $q = \text{—}$.

proposition N 24 in 14 targets; S 2 in 2 works; C 10; S/C rate ratio 0.83; panel $q = 0.858$.

overture N 7 in 7 targets; S 6 in 5 works; C 3; S/C rate ratio 8.27; panel $q = 0.055$.

insist N 23 in 16 targets; S 2 in 2 works; C 14; S/C rate ratio 0.59; panel $q = 0.641$.

mediation N 6 in 6 targets; S 2 in 2 works; C 4; S/C rate ratio 2.07; panel $q = 0.641$.

avow N 14 in 14 targets; S 2 in 2 works; C 10; S/C rate ratio 0.83; panel $q = 0.858$.

parliament N 61 in 24 targets; S 23 in 6 works; C 48; S/C rate ratio 1.98; panel $q = 0.477$.

passport N 18 in 16 targets; S 3 in 3 works; C 16; S/C rate ratio 0.78; panel $q = 0.798$.

memorial N 9 in 8 targets; S 4 in 3 works; C 10; S/C rate ratio 1.65; panel $q = 0.641$.

propose N 56 in 37 targets; S 22 in 16 works; C 62; S/C rate ratio 1.47; panel $q = 0.397$.

secretary N 56 in 37 targets; S 8 in 1 works; C 76; S/C rate ratio 0.44; panel $q = \text{—}$.

commission N 69 in 30 targets; S 54 in 22 works; C 110; S/C rate ratio 2.03; panel $q = 0.124$.

audience N 47 in 23 targets; S 33 in 16 works; C 77; S/C rate ratio 1.77; panel $q = 0.176$.

allowance N 23 in 16 targets; S 11 in 7 works; C 39; S/C rate ratio 1.17; panel $q = 0.798$.

N = Neville; S = Shakespeare; C = comparison drama. Panel q values correct the eligible individual tests within the twenty-one-term family. A dash means that the minimum token-and-dispersion rule was not met.

5.4 What the individual-word screen shows

Neville uses *advertisement* 70 times across 40 targets, or 515.6 per million tokens. Shakespeare uses it four times across three works, or 4.44 per million; comparison drama uses it seven times across six plays, or 1.88 per million. Shakespeare's rate is therefore 2.36 times the comparison rate, but the full-vocabulary corrected q is 0.52. Three of the four Shakespeare examples are assigned to 1597–1598, so the word does not show a post-1599 rise. Neville uses exact *proposition* 24 times across fourteen targets. Shakespeare uses it twice across two works. Its Shakespeare/comparison rate ratio is 0.83, with corrected $q = 0.94$. Its interest is chronological and contextual rather than an overall excess in Shakespeare. Ten lemmas from the declared broad diplomatic core are individually more prevalent in Shakespeare than in comparison drama after correction across the entire 8,215-test family:

- *answer*: Shakespeare/comparison rate ratio 1.65; corrected $q = < 0.00001$.
- *minister*: Shakespeare/comparison rate ratio 2.33; corrected $q = 0.00030$.
- *whereof*: Shakespeare/comparison rate ratio 2.18; corrected $q = 0.00042$.
- *demand*: Shakespeare/comparison rate ratio 2.29; corrected $q = 0.00094$.
- *conclusion*: Shakespeare/comparison rate ratio 2.55; corrected $q = 0.004$.
- *proceed*: Shakespeare/comparison rate ratio 1.52; corrected $q = 0.009$.
- *article*: Shakespeare/comparison rate ratio 2.76; corrected $q = 0.013$.
- *reply*: Shakespeare/comparison rate ratio 1.94; corrected $q = 0.022$.
- *overture*: Shakespeare/comparison rate ratio 8.27; corrected $q = 0.045$.
- *practice*: Shakespeare/comparison rate ratio 1.70; corrected $q = 0.047$.

Polysemous lemmas such as *article*, *practice*, and *minister* require concordance-level sense review. No individual diplomatic-core lemma passes $q < 0.05$ for the Shakespeare transition from before 1599 to 1599–1603. The chronology is strongest at the panel level.

5.5 Matched-panel calibration

Among 20,000 matched panels, the observed overall Shakespeare/comparison ratio of 1.54 has an empirical upper-tail probability of 0.036. The observed Shakespeare 1601–1603/pre-1599 ratio of 2.59 has probability 0.011. The 1601–1603 Shakespeare/comparison ratio of 2.64 has probability 0.171, because matched peak-period ratios are highly variable. Across matching neighborhoods of twenty to eighty terms and several seeds, the empirical probability ranges from 0.027 to 0.046 for overall Shakespeare elevation and from 0.008 to 0.012 for the internal Shakespeare chronological rise. Excluding *treaty*, the hardest high-frequency term to match, gives 0.039 and 0.0089 respectively. This supports the chronological pattern but remains a post hoc sensitivity analysis.

6 Interpretation

The evidence supports a pattern claim with four components: concentration in Neville, entry and growth in Shakespeare at a defined chronological break, clustering across a coordinated semantic register, and comparison against other drama. It does not require the constituent words to be rare in print. Common availability and unusual dramatic configuration are compatible facts. The study does not establish a causal mechanism. Direct influence, access to shared diplomatic discourse, collaboration, topical pressure, and

other transmission routes remain possible. Nor does a vocabulary-rate result establish authorship. What the analysis establishes is that the diplomatic panel’s Shakespeare prevalence changes in the predicted interval and that the peak is not mirrored by comparison drama.

7 Limitations

- Neville is represented mainly by correspondence, while Shakespeare and the control corpus are represented mainly by drama. Genre is an unavoidable confound.
- Composition years are estimates. The era bins are transparent analytical conventions, not exact transmission dates.
- Lemmatization controls spelling variation but can collapse distinct senses and introduce tagging errors.
- Stage directions are included in dramatic token streams. A speech-only sensitivity analysis would be useful.
- Shakespeare includes poems and a sonnet sequence as explicitly listed in Appendix A; the comparison corpus contains plays only.
- The two corpora are not disjoint: *Pericles* and *The Two Noble Kinsmen* appear in both. This introduces slight dependence between the Shakespeare and comparison estimates; a disjoint-corpus sensitivity analysis would be useful.
- The comparison corpus contributes to panel selection and evaluation, so a new held-out corpus would provide a stronger external test.
- Matched-panel calibration is post hoc and approximate; its result is supportive rather than confirmatory.

8 Conclusion

The systematic study changes the emphasis from memorable single words to an observable register. *Advertisement* is elevated descriptively but not individually reliable after correction. *Proposition* is chronologically suggestive but not more common in Shakespeare than in comparison drama overall. The twenty-one terms collectively tell a stronger story: Shakespeare’s rate rises from 153.1 per million before 1599 to 396.3 in 1601–1603, while comparison drama remains near 150. The corresponding rate ratio is 2.64, and the work-level uncertainty excludes parity. The most defensible conclusion is therefore that Shakespeare’s language becomes unusually saturated with a Neville-concentrated diplomatic-administrative vocabulary in 1599–1603. Historical explanation requires contextual reading and independent documentary evidence, but the statistical pattern itself is clear and reproducible.

A Exact Shakespeare works counted in each era

This appendix is the complete Shakespeare denominator. “Year used” is the composition-year assignment used in the calculations; “type” shows why the corpus contains forty-two works but is not a plays-only corpus. Public text access: [Folger Shakespeare Library](#).

Work	Year used	Type	Tokens
Before 1599: 20 works; 418,093 tokens			
The Two Gentlemen of Verona	1590	Play	17,077
Henry VI, Part 2	1591	Play	25,348
Henry VI, Part 3	1591	Play	24,145
The Taming of the Shrew	1591	Play	20,929
Henry VI, Part 1	1592	Play	21,439
Richard III	1592	Play	29,091
Titus Andronicus	1592	Play	20,516
Venus and Adonis	1593	Poem	9,947
Lucrece	1594	Poem	15,129
The Comedy of Errors	1594	Play	14,688
A Midsummer Night’s Dream	1595	Play	16,506
Love’s Labor’s Lost	1595	Play	21,163
Richard II	1595	Play	22,182
Romeo and Juliet	1595	Play	24,379
King John	1596	Play	20,775
The Merchant of Venice	1596	Play	21,267
Henry IV, Part 1	1597	Play	24,642
Henry IV, Part 2	1597	Play	26,210
The Merry Wives of Windsor	1597	Play	21,536
Much Ado About Nothing	1598	Play	21,124
1599-1600: 3 works; 67,455 tokens			
As You Like It	1599	Play	21,730
Henry V	1599	Play	26,146
Julius Caesar	1599	Play	19,579
1601-1603: 6 works; 121,110 tokens			
Hamlet	1601	Play	30,305
The Phoenix and Turtle	1601	Poem	358
Twelfth Night	1601	Play	19,851
Troilus and Cressida	1602	Play	25,983
All’s Well That Ends Well	1603	Play	23,026
Measure for Measure	1603	Play	21,587
1604-1615: 13 works; 294,102 tokens			
Othello	1604	Play	26,304
King Lear	1605	Play	25,845
Timon of Athens	1605	Collaborative play	18,294

Work	Year used	Type	Tokens
Antony and Cleopatra	1606	Play	24,682
Macbeth	1606	Play	17,127
Coriolanus	1608	Play	27,656
Pericles	1608	Collaborative play	18,220
Shakespeare's Sonnets	1609	Sonnet sequence	17,846
The Winter's Tale	1609	Play	24,963
Cymbeline	1610	Play	27,513
The Tempest	1611	Play	16,630
Henry VIII	1613	Collaborative play	24,712
The Two Noble Kinsmen	1613	Collaborative play	24,310

B Exact comparison-corpus plays counted in each era

This appendix is the complete comparison denominator: 203 plays dated 1590–1615, grouped by the same four era rules. Because the operational exclusion removes matched First Folio plays, the list retains *Pericles* and *The Two Noble Kinsmen*; both also appear in Appendix A as Shakespeare collaborations. Public corpus documentation: [Folger Early Modern English Drama](#).

Play	Year used	Tokens
Before 1599: 39 works; 634,764 tokens		
Antonius	1590	14,405
Arden of Faversham	1590	20,200
Fair Em	1590	12,042
Jack Straw	1590	8,201
Love's Metamorphosis	1590	10,297
The Love of David and Fair Bathsheba	1590	15,338
The Reign of King Edward the Third	1590	19,919
The Scottish History of James the Fourth	1590	20,661
The Taming of a Shrew	1590	12,762
The True Chronicle of King Leir	1590	21,778
1 Selimus	1591	18,865
1 The Troublesome Reign of King John	1591	14,384
2 The Troublesome Reign of King John	1591	9,883
Descensus Astraeae	1591	1,125
Edward the First	1591	22,629
Locrine	1591	16,368
Mucedorus (and Amadine)	1591	12,369
Orlando Furioso	1591	11,225
A Knack to Know a Knave	1592	16,691
Caesar and Pompey, or Caesar's Revenge	1592	19,382
Edward the Second	1592	21,157
Summer's Last Will and Testament	1592	16,940
The Massacre at Paris	1593	10,357
The Woman in the Moon	1593	12,664

Play	Year used	Tokens
Alphonsus, Emperor of Germany	1594	20,665
Cleopatra	1594	13,513
Cornelia	1594	15,009
The Four Prentices of London	1594	21,973
Two Lamentable Tragedies	1594	18,835
Mustapha	1596	13,607
The Blind Beggar of Alexandria	1596	13,278
1 The Fair Maid of the West	1597	15,015
An Humorous Day's Mirth	1597	16,545
The Case Is Altered	1597	18,230
Englishmen for My Money	1598	21,919
Every Man in His Humour	1598	24,937
Histrionmastix, or The Player Whipped	1598	13,508
The Death of Robert, Earl of Huntingdon	1598	22,644
The Two Angry Women of Abingdon	1598	25,444
1599-1600: 19 works; 393,359 tokens		
1 Edward the Fourth	1599	23,363
1 Sir John Oldcastle	1599	21,672
2 Edward the Fourth	1599	25,139
Antonio and Mellida	1599	13,474
Clyomon and Clamydes	1599	23,951
Every Man Out of His Humour	1599	37,978
Old Fortunatus	1599	23,864
The Merry Devil of Edmonton	1599	11,963
The Shoemaker's Holiday	1599	19,533
The Thracian Wonder	1599	17,569
The Trial of Chivalry	1599	19,561
Thomas Lord Cromwell	1599	13,727
Alaham	1600	21,666
Antonio's Revenge	1600	15,772
Cynthia's Revels	1600	32,749

Play	Year used	Tokens
Jack Drum's Entertainment	1600	16,730
Lust's Dominion, or The Lascivious Queen	1600	20,439
Patient Grissel	1600	20,130
The Maid's Metamorphosis	1600	14,079
1601-1603: 27 works; 472,918 tokens		
All Fools	1601	18,950
Blurt, Master Constable	1601	17,293
How a Man May Choose a Good Wife from a Bad	1601	21,656
Poetaster	1601	28,787
Satiromastix, or The Untrussing of the Humorous Poet	1601	22,495
The Contention between Liberality and Prodigality	1601	9,891
What You Will	1601	17,403
A Larum for London, or The Siege of Antwerp	1602	12,647
Hoffman, or A Revenge for a Father	1602	19,908
Mariam, the Fair Queen of Jewry	1602	17,975
May Day	1602	23,191
Sir Giles Goosecap	1602	20,651
The Fair Maid of the Exchange	1602	20,958
The Famous History of Sir Thomas Wyatt	1602	11,283
The Gentleman Usher	1602	21,113
The Royal King and the Loyal Subject	1602	19,857
The Spanish Tragedy [Additions]	1602	2,754
Wily Beguiled	1602	19,728
A Woman Killed with Kindness	1603	16,571
Darius	1603	17,262
King James his Royal and Magnificent Entertainment through his Honorable City of London	1603	7,257
Nobody and Somebody	1603	15,801
The Dutch Courtesan	1603	18,315
The Family of Love	1603	19,983
The London Prodigal	1603	16,157
The Malcontent	1603	15,061

Play	Year used	Tokens
The Phoenix	1603	19,971
1604-1615: 118 works; 2,224,464 tokens		
1 If You Know Not Me You Know Nobody	1604	12,141
1 Jeronimo, with the Wars of Portugal	1604	9,536
1 The Honest Whore	1604	23,991
A Mad World, My Masters	1604	18,271
Arches of Triumph	1604	6,799
Bussy D'Ambois	1604	19,906
Caesar	1604	25,276
Croesus	1604	22,765
Law Tricks	1604	15,623
Michaelmas Term	1604	19,827
Parasitaster, or The Fawn	1604	22,473
Sejanus His Fall	1604	25,642
The Entertainment of the Queen and Prince at Althorp (The Satyr)	1604	5,530
The Vision of the Twelve Goddesses	1604	1,724
The Widow's Tears	1604	21,572
The Wise Woman of Hogsdon	1604	18,330
Westward Ho	1604	22,805
When You See Me You Know Me	1604	24,351
2 If You Know Not Me You Know Nobody	1605	22,070
2 The Honest Whore	1605	23,678
A Trick to Catch the Old One	1605	18,431
A Yorkshire Tragedy	1605	5,989
Caesar and Pompey (The Wars of Caesar and Pompey)	1605	17,267
Eastward Ho	1605	24,287
Monsieur D'Olive	1605	17,559
Northward Ho	1605	21,190
Philotas	1605	16,861
The Alexandraean Tragedy	1605	26,683
The Characters of Two Royal Masques	1605	7,169

Play	Year used	Tokens
The Miseries of Enforced Marriage	1605	24,082
The Noble Gentleman	1605	19,199
The Queen's Arcadia	1605	20,341
The Revenger's Tragedy	1605	20,161
The Triumphs of Reunited Britannia	1605	4,434
The Wonder of Women, or Sophonisba	1605	13,499
Volpone	1605	26,821
Hymenaei	1606	9,982
Love's Cure, or The Martial Maid	1606	20,195
The Devil's Charter	1606	24,310
The Fleer	1606	16,195
The Isle of Gulls	1606	21,352
The Puritan, or The Widow of Watling Street	1606	18,910
The Rape of Lucrece	1606	20,370
The Two Maids of More-Clacke	1606	16,845
The Woman Hater	1606	22,675
A Shoemaker a Gentleman	1607	20,668
Cupid's Revenge	1607	20,868
Cupid's Whirligig	1607	23,200
Humor Out of Breath	1607	14,166
The Conspiracy of Charles Duke of Byron	1607	16,452
The Dumb Knight	1607	18,915
The Insatiate Countess	1607	20,814
The Knight of the Burning Pestle	1607	21,360
The Masque at Lord Hay's Marriage	1607	5,447
The Tragedy of Charles Duke of Byron	1607	18,756
The Travels of the Three English Brothers	1607	16,389
The Turk	1607	17,639
Your Five Gallants	1607	20,359
Pericles	1608	18,220
Ram Alley	1608	20,977

Play	Year used	Tokens
The Birth of Merlin	1608	19,294
The Bloody Banquet	1608	15,102
The Coxcomb	1608	21,179
The Faithful Shepherdess	1608	20,361
A Christian Turned Turk	1609	19,129
A Woman Is a Weathercock	1609	16,728
Epicoene or The Silent Woman	1609	26,123
Fortune by Land and Sea	1609	18,211
Philaster	1609	20,061
The Captain	1609	24,578
Amends for Ladies	1610	18,270
Chester's Triumph	1610	3,431
The Alchemist	1610	14,743
The Golden Age, or The Lives of Jupiter and Saturn	1610	18,844
The Maid's Tragedy	1610	21,345
The Revenge of Bussy D'Ambois	1610	20,004
The Scornful Lady	1610	23,155
Valentinian	1610	24,970
A King and No King	1611	24,259
A New Wonder, A Woman Never Vexed	1611	22,465
Bonduca	1611	20,778
Catiline His Conspiracy	1611	18,905
Fuimus Troes	1611	17,486
Greene's Tu Quoque, or The City Gallant	1611	23,551
If It Be Not Good, the Devil Is in It	1611	22,098
Match Me in London	1611	18,221
The Atheist's Tragedy, or The Honest Man's Revenge	1611	21,079
The Brazen Age	1611	21,066
The Night Walker, or The Little Thief	1611	19,996
The Roaring Girl, or Moll Cutpurse	1611	24,048
The Woman's Prize, or The Tamer Tamed	1611	23,243

Play	Year used	Tokens
1 The Iron Age	1612	18,940
2 The Iron Age	1612	17,408
The White Devil	1612	25,210
A Chaste Maid in Cheapside	1613	17,274
Cynthia's Revenge	1613	32,919
Four Plays, or Moral Representations, in One	1613	21,239
The Hog Hath Lost His Pearl	1613	16,718
The Honest Man's Fortune	1613	23,542
The Masque of the Inner Temple and Gray's Inn	1613	2,891
The Masque of the Middle Temple and Lincoln's Inn	1613	4,297
The Triumphs of Truth; The Entertainment on Michaelmas Day 1613	1613	6,415
The Two Noble Kinsmen	1613	24,310
Wit at Several Weapons	1613	23,030
Bartholomew Fair	1614	36,388
Hymen's Triumph	1614	15,092
More Dissemblers Besides Women	1614	19,198
The Duchess of Malfi	1614	23,745
The Hector of Germany, or The Palsgrave, Prime Elector	1614	16,147
The Honest Lawyer	1614	21,150
The Sophister	1614	15,931
The Staple of News	1614	25,439
A Fair Quarrel	1615	20,216
Albumazar	1615	21,714
Band, Cuff, and Ruff, or Exchange Ware at the Second Hand	1615	1,631
Beggars' Bush	1615	19,291
Sicelides	1615	21,133
The Widow	1615	19,126

C Reproducibility specification

Item	Specification
Neville units / tokens	164 letters plus three supplementary documents; 135,770 tokens
Shakespeare units / tokens	42 whole works; 900,760 tokens
Comparison units / tokens	203 whole plays; 3,725,505 tokens
Lexical inventory	16,850 lower-case lemmas after declared function-word and malformed-token exclusions
Word-test eligibility	At least two tokens across at least two whole units in each compared corpus
Uncertainty	Whole-unit sandwich variance; 95% normal intervals
Multiplicity	Benjamini–Hochberg adjustment within each comparison family
Panel bootstrap	20,000 whole-work replicates; seed 1599
Matched calibration	20,000 panels; Shakespeare-blind matching; primary neighborhood 40; seed 1599
Zero cells	Half-token descriptive correction; no formal p or q

References

- [1] Yoav Benjamini and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society, Series B* 57.1 (1995): 289–300. DOI metadata.
- [2] Folger Shakespeare Library. [Shakespeare’s Works: Downloadable Texts](#). Accessed 31 July 2026.
- [3] Folger Shakespeare Library. [Early Modern English Drama](#). Accessed 31 July 2026.
- [4] EarlyPrint. [Early English print corpus and research platform](#). Accessed 31 July 2026.
- [5] Text Creation Partnership. [Public project site](#). Accessed 31 July 2026.
- [6] The National Archives. [State Papers Foreign, 1509–1782](#). Accessed 31 July 2026.
- [7] Ralph Winwood. *Memorials of Affairs of State in the Reigns of Queen Elizabeth and King James I.* 1725. [Internet Archive scan](#).
- [8] Tove Larsson and Gregory R. Hancock. *Statistical Modeling for (Actual) Hypothesis Testing: Building Cumulative Knowledge in Corpus Linguistics*. Cambridge University Press, 2026. DOI 10.1017/9781009660877.