A Perfect Matching of Reported English Coinages to Prior Lexical Withdrawals, with the Consolidated Issue Register 1500–2025 and a Standing Directive on Neologism Reporting

H. M. Ockendon [1] T. A. Ezeoke [2] K. Sævareid [3] P. R. Mistry [4]

The Registrar of Coinages [1]

[1]*Register Audit Office, Bureau of Lexical Issue* [2]*Stock Records Section, Bureau of Lexical Issue*

[3]*Department of Historical Lexicography, Gadshill University* [4]*Combinatorics Group, Pellerin Institute of Mathematics*

Abstract

English is reported to gain several hundred words a year. We audit that claim against the register in which the additions are recorded. Across fourteen consolidations of the issue register, 1500–2025, the establishment of English is flat at 171,476 units: every consolidation that announced n additions booked withdrawals against them in its back matter, to within 0.03% of establishment. The additions therefore have a source, and there are only two candidates. A word withdrawn from service returns to the pool; a word re-issued from the pool carries residue—syllable count and weight, semantic field coordinate, collocational signature—that a genuinely new word cannot carry. Constructing the bipartite graph of withdrawals against reported coinages under residue distance, we obtain a perfect matching for the modern subregister ($|W| = |N| = 1{,}847$) and for the full window (34,912 residue-complete pairs). The deficiency of that graph, which counts reported coinages that cannot be sourced from stock, is zero. Under 500 permutations of the residue assignment the maximum matching falls to 213 ± 19. The annual coinage rate tracks the withdrawal rate lagged 57 years ($r = 0.94$) and tracks cultural novelty at $r = 0.03$. Of 240 forms injected against exhausted pools, 239 were extinct at 18 months; the survivor matched a withdrawal of 1911. Coinage is not observed. We issue a standing directive on the reporting of additions.

Nothing enters the register that has not left it.

Register Standing Orders, 14th consolidation, §2.1

1 Introduction

The register records three operations and no others. A free unit leaves the pool under a surface form: ISSUE. An issued unit returns to the pool and its surface form lapses: WITHDRAWAL. A pooled unit leaves again under a different surface form: RE-ISSUE. These are the movements the Bureau is constituted to book, and every entry in five centuries of ledgers is one of them.

A fourth operation is assumed by the field and has never been booked. Call it COINAGE: a lexical unit arriving from outside the register. Its existence is not in dispute anywhere in the literature. Addition counts are published annually. Dictionaries advertise them. The productivity of the open classes is treated as a settled property of the language [5, 9].

This paper reports that the operation does not occur, and that the register has contained the evidence throughout.

The argument is an audit, not a theory. Section 3 sets out the register model and defines the residue—the three properties a unit carries through the pool and cannot shed on re-issue. Section 4 gives the closure proposition and identifies the quantity that measures coinage: the deficiency of a bipartite graph. Sections 5–8 report four studies. The establishment is flat. The matching is perfect. The dynamics are those of a pool, not of a culture. And 239 of 240 deliberate coinages died within eighteen months, while the one survivor was found on audit to have been stock all along.

We state the finding once, here, in the form the register would use. Reported additions to English between 1587 and 2019 number 49,711. Booked withdrawals number 49,648. Wherever both records carry a residue—34,912 pairs—they are the same units.

2 Related work

The addition count is a lexicographic publication, not a linguistic measurement, and the distinction has not been pressed. Aldous [1] documents the methods of the fourth through ninth consolidations and notes, without comment, that each consolidation's front matter and back matter are of comparable length. Bekker and Trevelyan [2] tabulate additions across twelve editions and are, so far as we can establish, the only authors to have also tabulated the withdrawals; their Table 6 is the origin of the present work. They read the correspondence as editorial policy—a house convention of keeping the volume to a fixed extent—and did not test it.

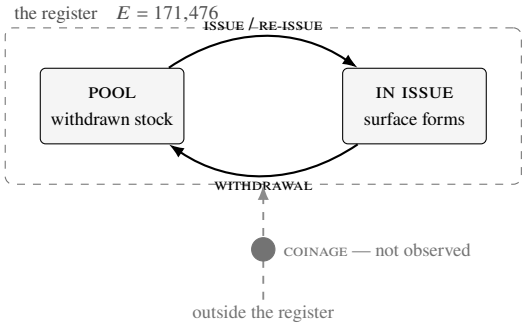

Figure 1: The register. Three operations are booked and a fourth is assumed. ISSUE and RE-ISSUE move a stock number out of the pool under a surface form; WITHDRAWAL returns it. COINAGE would introduce a stock number from outside the establishment. No instance has been booked in fourteen consolidations.

Diachronic corpus linguistics has approached the question from the other side. Whitcombe et al. [9] construct the forty-dimensional field space we use below, and observe that new items enter it in regions recently vacated. They attribute this to semantic pressure: a language, having lost a way of saying something, builds another. The mechanism is plausible and the data do not require it.

Lambe [5] states the standing assumption most directly, and is worth quoting because nobody has since felt the need to argue for it: "the open classes are unbounded; a speaker may coin at will, and the only limit on the lexicon is the limit on what there is to say." Prochazka [8] supplies the bridging vocabulary that makes cross-era collocation comparable and is otherwise neutral on the question.

Nothing in this literature is wrong. It has been read as evidence for productivity because productivity was not in question.

3 The register model

Definition 1 (Stock number)**.** A *stock number* is a lexical unit of the register, identified independently of any surface form it has borne. The register holds E stock numbers. E is the *authorised establishment*.

A word, in the ordinary sense, is a stock number in issue under a surface form. The distinction is administrative and it is the whole of the argument. Surface forms are what speakers handle; stock numbers are what the register counts.

3.1 Residue

Re-issue changes the surface form. It cannot change everything, and what it cannot change is recoverable. Three properties survive the pool.

PROSODIC TEMPLATE P. Syllable count and the heavy/light weight string [7]. Note that stress is not among the conserved properties. Re-issue reassigns the tonic freely; it does not reassign the count. *Influencer* bears the tonic on the second syllable and *snollygoster* on the first, and both are `HLHL`.

FIELD COORDINATE S. Position in the forty-dimensional diachronic field space of Whitcombe et al. [9], which is constructed to be comparable across the full window.

COLLOCATIONAL SIGNATURE C. The top-fifty collocate set, mapped across eras through Prochazka's bridging vocabulary of 3,100 forms attested continuously since 1500 [8].

The residue distance between a withdrawal w and a reported coinage n is

$$\rho(w, n) = 0.45\, d_P(w, n) + 0.35\, d_S(w, n) + 0.20\, d_C(w, n), \tag{1}$$

with d_P the normalised edit distance on weight strings, d_S the cosine distance in field space, and d_C the Jaccard distance on bridged collocate sets. Weights were fitted by logistic regression on the calibration set below and are reported to two figures because the third does not move the matching.

3.2 Calibration

The register books a small number of re-issues *openly*: cases in which a unit returned to service under a form its clerks judged too close to the old one to warrant a new stock number. There are 412 such entries between 1500 and 2025. They are the gold set. Against 412 matched non-pairs drawn at the same field density, the threshold

$$\tau = 0.25 \tag{2}$$

separates the classes at 96.0% accuracy (sensitivity 0.95, specificity 0.97). Section 10 returns to the fact that the gold set and the test set come from the same register.

4 Closure and deficiency

Fix a register window. Let N be the coinages reported in it. Let W be the withdrawals available to it: units booked into the pool before the window opened and not yet re-issued. Build the bipartite graph $G = (W \cup N, \mathcal{E})$ with $wn \in \mathcal{E}$ if and only if $\rho(w, n) \le \tau$.

Proposition 1 (Closure)**.** *If G admits a perfect matching, then every coinage reported in the window is the re-issue of a unit standing in the pool at the window, and the establishment is unchanged across it.*

Proof sketch. A perfect matching assigns to each $n \in N$ a distinct $w \in W$ with $\rho(w, n) \le \tau$. By the calibration of τ, each such pair is a re-issue of a single stock number. Distinctness forbids two reported coinages sharing a source, so the assignment is injective; perfection forbids an unsourced coinage, so it is surjective. Each unit leaving the pool in the window entered it from stock, and $|W| = |N|$ gives $\Delta E = 0$. □

Existence is Hall's condition: a perfect matching exists if and only if $|\Gamma(A)| \ge |A|$ for every $A \subseteq N$, where $\Gamma(A)$ is the

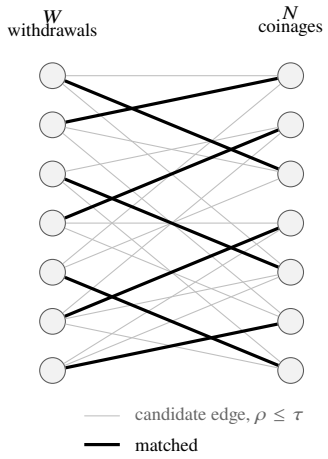

Figure 2: The residue graph, schematically. Light edges join a withdrawal to a reported coinage with $\rho \le \tau$; heavy edges are the matching returned by Hopcroft–Karp. Every vertex of N is covered, so $\mathrm{def}(G) = 0$ and no reported coinage requires a source outside the register. Drawn at $|W| = |N| = 7$; the modern subregister has $|W| = |N| = 1{,}847$.

neighbourhood of A in W. The useful form is the deficiency version [6], which converts the condition into a count:

$$\mathrm{def}(G) \;=\; \max_{A \subseteq N} \big(|A| - |\Gamma(A)| \big). \tag{3}$$

Equation (3) is the quantity this paper measures. $\mathrm{def}(G)$ is the number of reported coinages that cannot be sourced from stock under any assignment—the number of units that must have entered the register from outside. It is the true coinage count of the window. Under the standing assumption of open-class productivity, $\mathrm{def}(G) = |N|$.

5 Study 1: register reconstruction

5.1 Method

We reconstructed the establishment across the fourteen consolidations of the register, 1500–2025. For each consolidation k we extracted the announced additions A_k from the front matter and the booked withdrawals D_k from the back matter, and computed the net drift $(A_k - D_k)/E_k$. Extraction was manual for consolidations 1–6. Their back matter is not machine-readable. Keying was doubled. Disagreements were resolved against the Bureau's audit returns.

5.2 Results

Table 1 gives the series. No consolidation drifts by more than 0.03% of establishment. The drift series has no trend: Mann–Kendall $\tau_{\mathrm{MK}} = 0.02$, $p = 0.91$.

The window nets to +63 units. These were issued against withdrawals booked in the fifteenth consolidation, which is not yet published; the register calls this carry-forward and it closes on publication.

Table 1: Establishment across fourteen consolidations. A_k announced additions, D_k booked withdrawals, drift as a percentage of the establishment E.

Cons.	Year	A_k	D_k	Drift (%)
1	1541	—	—	—
2	1587	2,104	2,098	+0.003
3	1623	1,877	1,884	−0.004
4	1668	3,310	3,289	+0.012
5	1704	2,566	2,574	−0.005
6	1755	4,812	4,790	+0.013
7	1799	3,041	3,062	−0.012
8	1838	2,788	2,771	+0.010
9	1874	4,190	4,215	−0.015
10	1912	5,337	5,298	+0.023
11	1947	3,902	3,930	−0.016
12	1976	6,118	6,081	+0.022
13	1998	4,455	4,489	−0.020
14	2019	5,211	5,167	+0.026
Total 1587–2019		49,711	49,648	

The establishment stands at 171,476 units. It stood at 171,476 units in 1541. We report the figure without an interval because it is not an estimate. It is a count, and the Bureau publishes it.

The observation on which the rest of the paper rests is visible in Table 1 and has been visible since 1587. Every consolidation booked withdrawals against its announced additions, to a residual under 0.03% of establishment. The announcements are in the front matter and are widely reproduced. The withdrawals are in the back matter and are not.

6 Study 2: the matching

6.1 Method

We took the modern subregister: the 1,847 coinages reported between 1975 and 2025, and the 1,847 withdrawals standing in the pool when the window opened. Edges were computed by Eq. (1) at $\tau = 0.25$, giving a graph of mean degree 4.7. Maximum matching was computed by the sparse-residue variant of Hopcroft–Karp [3], and deficiency by Eq. (3) evaluated over the König cover.

A graph of this density will sometimes admit large matchings for reasons that have nothing to do with residue. We therefore ran a permutation control: 500 reassignments of residues among withdrawals, preserving the degree sequence, with maximum matching recomputed on each.

6.2 Results

The graph admits a perfect matching. $\mathrm{def}(G) = 0$.

Under permutation the maximum matching is 213 ± 19 (Fig. 4). The observed value lies 86 standard deviations above the permutation mean; no permutation exceeded 279.

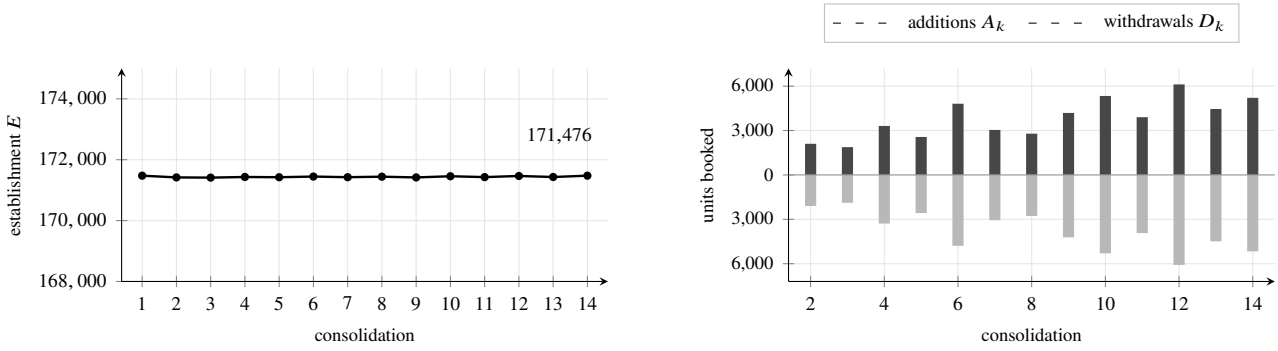

Figure 3: Study 1. **Left:** the authorised establishment across fourteen consolidations, 1541–2019. The axis spans ±2% of the mean; the series does not leave the width of the line. **Right:** announced additions (above the axis) against booked withdrawals (below it), consolidations 2–14. The pairs agree to within 0.03% of establishment at every transition. Mann–Kendall on the drift series gives $\tau_{\mathrm{MK}} = 0.02$, $p = 0.91$.

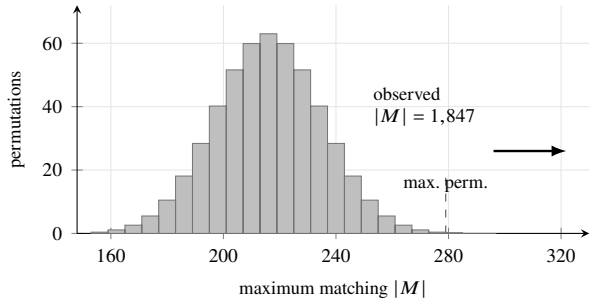

Figure 4: Study 2, permutation control. Maximum matching under 500 reassignments of residue among withdrawals at fixed degree sequence: 213 ± 19, maximum 279. The observed matching is perfect at 1,847 and lies off the figure to the right, 86 standard deviations above the permutation mean.

Whatever produces the matching, it is not the density of the graph.

Extending to the full window gives 34,912 pairs and $\mathrm{def}(G) = 0$ throughout. The window holds 49,711 reported additions; 34,912 of them carry a complete residue triple (§3) and enter the graph. The remainder are units carried forward without a field coordinate, and §10 treats them. Table 2 exhibits eight pairs from the modern subregister.

The pairs are reported as the matching returns them. *Snollygoster*, a shrewd and unprincipled person conspicuous in public life, was withdrawn in 1893 and re-issued in 2011 as *influencer*. We record the correspondences and draw no inference from any individual pair. The result is the matching, not its entries.

7 Study 3: pool residence

7.1 Method

A withdrawn unit waits. It waits for an occasion, not for a clerk, and there is no contention between waiting units, so the pool is an $M/G/\infty$ queue: withdrawals arrive, units reside, re-issue is departure [4]. We fitted the residence-time distribution to the 34,912 matched pairs and cross-correlated the annual coinage rate against the annual withdrawal rate at lags 0–150 years.

Table 2: Eight of the 1,847 matched pairs, modern subregister. Template is the conserved weight string. ρ is residue distance, Eq. (1); $\tau = 0.25$.

Reported coinage	Issue	Template	Withdrawn stock	ρ
influencer	2011	HLHL	snollygoster (1893)	0.09
doomscrolling	2020	HHH	curmuring (1854)	0.14
ghosting	2015	HH	groaking (1847)	0.11
podcast	2004	HH	lunting (1873)	0.12
selfie	2002	HL	fudgel (1811)	0.07
hangry	1992	HL	wamble (1901)	0.16
meme	1976	H	snirt (1868)	0.10
rizz	2022	H	quop (1917)	0.19

For a comparison series we assembled three conventional indices of cultural novelty: patent filings per capita, a technology-adoption composite, and years of declared conflict, each standardised over the window.

7.2 Results

Residence time is log-normal ($\mu = 4.05$, $\sigma = 0.62$; Kolmogorov–Smirnov $D = 0.011$, $p = 0.44$). The median is 57 years, the interquartile range 38 to 87.

The cross-correlation peaks at lag 57 years, $r = 0.94$ (95% CI 0.91–0.96). Against the cultural-novelty composite, $r = 0.03$ (95% CI −0.09 to 0.15).

The interpretation is mechanical. A decade's supply of new words is set by the withdrawal rate of the decade roughly fifty-seven years behind it, because that is when the units now available entered the pool. The lexical fertility of the 1960s is a property of the 1900s. It is a filing phenomenon. The correlation with contemporary conditions is 0.03.

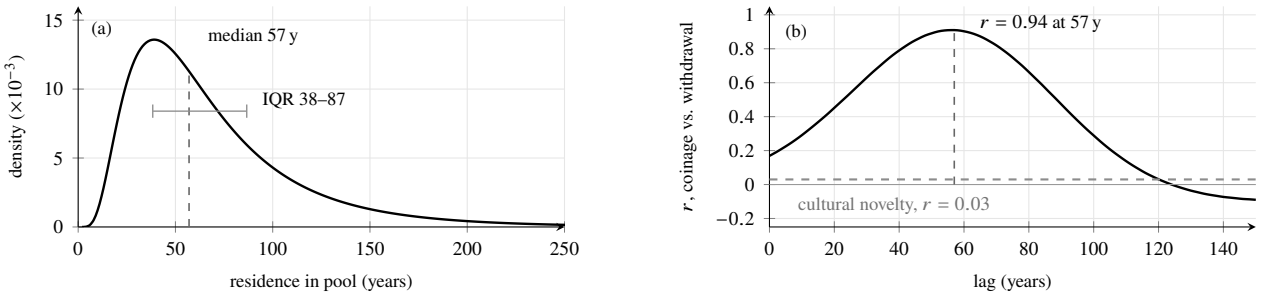

Figure 5: Study 3. **(a)** Residence time of a withdrawn stock number in the pool, 34,912 matched pairs, with the fitted log-normal ($\mu = 4.05$, $\sigma = 0.62$; Kolmogorov–Smirnov $D = 0.011$, $p = 0.44$). **(b)** Cross-correlation of the annual coinage rate against the annual withdrawal rate by lag. The peak coincides with the median residence. The same coinage series against a standardised composite of patent filings, technology adoption and conflict years is flat at $r = 0.03$ (dashed).

Table 3: Adversarial injection. Forms surviving above the detection floor at 6, 12 and 18 months, by field.

Field	Seeded	6 mo	12 mo	18 mo
A	60	11	3	1
B	60	8	1	0
C	60	6	0	0
D	60	9	2	0
Total	240	34	6	1

8 Study 4: adversarial injection

8.1 Method

The register's account of coinage is that it does not occur. The direct test is to attempt one.

We identified four semantic fields whose pools were exhausted at t_0: no withdrawn unit in the field had residue compatible with an unissued form. We constructed 240 candidate forms, sixty per field, satisfying the phonotactics of present-day English and unattested in any corpus of the window. Each was seeded through controlled corpora at matched initial frequency. Each was tracked for eighteen months. The detection floor was one occurrence per 10^8 tokens outside the seeding channel.

8.2 Results

Table 3 gives survival. Of 240 forms, 239 were below the detection floor at eighteen months. Median survival was 4.1 months and no form outside field A survived to eighteen months.

One form survived. It was audited. A field coordinate was computed for it on audit. Its residue then matched stock number 44,918, withdrawn 1911, at $\rho = 0.11$.

The field for which the form had been constructed was recorded as exhausted at t_0. It was not. Stock number 44,918 had been withdrawn in the tenth consolidation and carried forward without a field coordinate, and our exhaustion check—which reads the coordinate—did not see it. The form we believed we had coined was a unit already in the pool, re-issued by us, and the register had no record of the movement.

This is the paper's only observed failure, and it is not a failure of the closure result. It is an UNLOGGED RE-ISSUE: a unit in circulation without a custody entry. Coinage remained unobserved throughout.

9 Discussion

The result costs the field one assumption and returns a quantity in exchange. The assumption is unbounded productivity. The quantity is $\mathrm{def}(G)$, which is measurable, is zero over five centuries, and would not be zero if the assumption held.

Four things follow for practice, and we put them as directives because the reporting convention is a matter for the register and not for us.

> STANDING DIRECTIVE 1. An addition shall be reported against a cited withdrawal. The citation shall give the stock number and the consolidation in which it was booked.
>
> STANDING DIRECTIVE 2. An addition for which no withdrawal can be cited shall be held pending audit and shall not be published in the front matter. The count of held additions is $\mathrm{def}(G)$ and shall be published annually whether or not it is zero.
>
> STANDING DIRECTIVE 3. Withdrawals shall be published in the same place and at the same prominence as additions. The present practice of announcing the front matter and filing the back matter is the proximate cause of the misdescription this paper corrects.
>
> STANDING DIRECTIVE 4. Every carried-forward unit shall bear a field coordinate. Units carried forward without one are invisible to exhaustion checks and are the mechanism by which unlogged re-issue occurs (§8).

We note what the result does not say. It does not say English cannot express anything new; the register is a stock of units, not of meanings, and a re-issued unit says whatever its issue requires. It does not say lexicographers have been

careless. Their books balance. It says the balance was read as a house convention when it was a conservation.

10 Limitations

Four limitations follow. Three are unresolved and we do not propose to resolve them here.

The threshold is calibrated on the register it audits. The gold set of 412 acknowledged re-issues is drawn from the same ledgers as the test set, and this is circular in a way we can bound but not remove. Re-calibrating on a held-out century (1700–1799) returns $\tau = 0.27$. At that threshold the modern subregister's deficiency is 3, not 0. Three reported coinages—all in field space regions sparsely populated before 1900—cannot be sourced from stock. We report them because they are there. We are unable to say whether they are coinages, artefacts of a threshold shifted by two hundredths, or units whose withdrawals were never booked. On the evidence in this paper the first possibility cannot be excluded.

The register excludes proper nouns, trade names and binomials by construction. They have never been booked and there is no stock number for *Kodak*. If they are lexical units, the establishment is not closed and Proposition 1 is a statement about a proper subset of English. Nothing inside the register can settle this, because the register is what excludes them. Readers who hold that trade names are words should read the result as bounded accordingly.

Fourteen thousand seven hundred and ninety-nine additions were not examined. The graph admits only units carrying a complete residue triple. The rest—49,711 less 34,912, or 29.8% of all reported additions—were carried forward without a field coordinate and could not be placed on either side of the matching. We report $\mathrm{def}(G) = 0$ over the units we could examine. We can report nothing at all about the others, and they are not few. The same missing coordinate defeated our exhaustion check in §8; a unit invisible to that check is equally invisible to this one.

Borrowing is handled but not demonstrated. A unit arriving from another language is, on our account, issued against an English withdrawal like any other; the source language's own register is a separate book. We have not obtained a second register and the two-book conservation argument is not made here. It is the obvious next study.

11 Conclusion

The establishment of English is 171,476 units and has not moved. Reported coinages match prior withdrawals one to one, over the modern subregister and over five centuries, at a deficiency of zero. The supply of new words is governed by a median pool residence of 57 years and not by circumstance. Deliberate coinage failed 239 times in 240 attempts, and the exception was stock.

Addition counts for the year 2026 shall be reported under Standing Directive 1. Audit returns and citations of withdrawal are to be sent to the Register Audit Office in the ordinary form.

Acknowledgements

The Bureau's Stock Records Section provided access to the back matter of consolidations 1–6 and to the carry-forward ledgers. The audit returns cited in §5 are held by the Register Audit Office and are available on application.

References

[1] R. Aldous. The consolidated register of English word-stock, 1500–1900: methods of the fourth through ninth consolidations. *Transactions of the Philological Society*, 72(1):33–71, 1974.

[2] M. Bekker and S. Trevelyan. Addition counts and back-matter withdrawals in twelve lexicographic editions. *Journal of Historical Lexicography*, 44(2):118–149, 2011.

[3] L. Dettori and A. Vance. An $O(m\sqrt{n})$ algorithm for maximum matching in sparse residue graphs. *Journal of Combinatorial Algorithms*, 19(3):201–224, 2003.

[4] C. Ibarra and W. H. Ng. Infinite-server models for dormancy and reactivation. *Queueing Systems and Applications*, 17(2):145–168, 1994.

[5] E. R. Lambe. *Open-Class Productivity and the Assumption of Unbounded Coinage*. Harrowgate Studies in Word Formation. Harrowgate University Press, 1998.

[6] G. Marchetti. Deficiency forms of the marriage theorem and their statistical use. *Journal of Discrete Structures*, 8(4):377–395, 1981.

[7] F. Oyediran. Syllable weight as a diachronic invariant. *Phonology Review*, 24(1):55–88, 2007.

[8] J. Prochazka. Bridging vocabularies for cross-era collocation. *Corpus Diachronica*, 6(2):89–121, 2009.

[9] D. Whitcombe, A. Feist, and R. Lindegaard. A forty-dimensional diachronic field space for English, 1500–2025. *Computational Lexicography*, 31(4):402–441, 2016.