

Exact Enumeration of Word Trees with a Singleton Letter

Alper Ferudun*

27 October 2026

Abstract

Assign independent uniform maps of $[n]$ to the letters of a word. A word tree is an assignment for which the composite map has exactly one cyclic point. For every word containing a letter exactly once, we reduce its tree count to a two-point collision probability for the remaining word. A finite recursion on equality patterns of two trajectories evaluates this probability exactly; its state space depends on the word and not on n . In particular, for two maps the numbers of AB -trees and AAB -trees are $(2n-1)n^{2n-2}$ and $(3n^3 - 3n^2 - n + 2)n^{2n-4}$, respectively, for $n \geq 2$. These answer two explicitly proposed special cases of Chapuy's enumeration question. The proof uses the classical weighted Cayley identity and the conditional random-mapping construction already used by Martinsson. The general enumeration problem without a singleton letter is not resolved.

1 The question and the main formula

Write $[n] = \{1, \dots, n\}$, and let $\text{cyc}(f) = \{x : f^j(x) = x \text{ for some } j \geq 1\}$ for a map $f : [n] \rightarrow [n]$. The condition $|\text{cyc}(f)| = 1$ means that its functional digraph is a rooted tree with a loop at the root. It is stronger than having one connected component, whose cycle could have length greater than one.

Chapuy and Perarnau introduced word trees in their work on random automata [3]. Chapuy's Oberwolfach question [1, pp. 145–146] asks for exact formulas, recurrences, or functional equations, explicitly singling out $W = AB$. His 2022 lecture [2, slide 48] likewise proposes AAB as a starting case. We give closed formulas for both and a finite recursion for every word with a singleton letter.

We use the composition convention of the Oberwolfach report: if $W = L_1 \cdots L_k$, its map is $f_W = f_{L_1} \circ \cdots \circ f_{L_k}$. Thus trajectories read the letters from right to left. Using the opposite convention reverses words and does not change our $A^r B$ conclusions. Each distinct letter receives one uniform random map, independently; repeated occurrences use that same map.

Let V be a nonempty word on q distinct letters, and let B be a new letter. Put $g = f_V$, and define, for $n \geq 2$,

$$c_V(n) = \mathbb{P}(g(1) = g(2)).$$

Let $T_{VB}(n)$ count all assignments of the $q+1$ maps for which $g \circ f_B$ has exactly one cyclic point.

Theorem 1. *For every $n \geq 2$,*

$$T_{VB}(n) = n^{(q+1)n} \left(\frac{1}{n} + \frac{n-1}{n} c_V(n) \right). \quad (1)$$

There is an explicitly defined finite collection of nonnegative integers $d_V(v, e)$, depending only on V , such that

$$c_V(n) = \sum_{v,e} d_V(v, e) \frac{(n-2)_{v-2}}{n^e}. \quad (2)$$

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: [AlperTheKing](https://github.com/AlperTheKing).

The coefficients are produced by the recursion in Section 3. They vanish unless $2 \leq v \leq 2|V| + 2$ and $0 \leq e \leq 2|V|$. Consequently the probability $T_{VB}(n)/n^{(q+1)n}$ is a Laurent polynomial in n with integer coefficients. For $n = 1$, the count is 1.

Here $(x)_{\underline{m}} = x(x-1)\cdots(x-m+1)$, with $(x)_0 = 1$. If $v > n \geq 2$, the numerator in (2) is zero. The theorem does not enumerate all $n^{(q+1)n}$ map assignments: for a fixed word, its coefficients are computed once and then evaluated at any n . No polynomial bound in the word length is asserted.

Corollary 2. *For $n \geq 2$, the numbers of ordered pairs of maps are*

$$T_{AB}(n) = (2n-1)n^{2n-2}, \quad (3)$$

$$T_{AAB}(n) = (3n^3 - 3n^2 - n + 2)n^{2n-4}. \quad (4)$$

More generally, (1)–(2) give an exact finite formula for every $A^r B$, $r \geq 1$. The same counts apply to cyclic rotations and positive concatenation powers of these words.

The contribution is the exact enumeration in these explicitly specified cases and the finite coefficient recipe. The weighted mapping identity below is classical, and the conditioning step for $a^r \circ b$ appears in Martinsson’s proof [4, Section 3]. Our use of that step is to compute the event of exactly one cyclic point.

2 Weighted mappings and the collision reduction

Lemma 3 (Classical weighted Cayley consequence). *Suppose that $F(x)$, $x \in [n]$, are independent with common distribution $p = (p_1, \dots, p_n)$. Then*

$$\mathbb{P}(|\text{cyc}(F)| = 1) = \sum_{i=1}^n p_i^2. \quad (5)$$

Proof. For $n \geq 2$, fix a root i . The weighted Cayley identity for trees oriented towards i is

$$\sum_{T \text{ rooted at } i} \prod_{x \neq i} p_{\text{parent}_T(x)} = p_i \left(\sum_j p_j \right)^{n-2} = p_i.$$

For example, the usual Prüfer identity for undirected trees gives this after multiplying each tree weight by p_i : the exponent of p_i becomes its degree and that of every other vertex becomes its degree minus one. Requiring the loop $F(i) = i$ contributes another factor p_i . The possible roots are disjoint events, giving (5). The polynomial identity also permits zero probabilities. The case $n = 1$ is immediate. This is also the singleton-set case of Pitman’s cyclic-set formula [5, Eq. (27)]. $\square$

Proof of (1). First fix any deterministic $g : [n] \rightarrow [n]$, and set $s_i = |g^{-1}(i)|$. For an independent uniform map b , the values $g(b(x))$ are independent with common probabilities $p_i = s_i/n$. Lemma 3 therefore gives

$$\#\{b : |\text{cyc}(g \circ b)| = 1\} = n^{n-2} \sum_i s_i^2. \quad (6)$$

Now let $g = f_V$ be random. Every f_L has a conjugation-invariant law, and their joint law is invariant under simultaneous relabeling. Thus the probability that $g(x) = g(y)$ is $c_V(n)$ for any distinct x, y . Since

$$\sum_i s_i^2 = \sum_{x, y \in [n]} \mathbf{1}_{\{g(x)=g(y)\}},$$

its expectation is $n + n(n-1)c_V(n)$. There are n^{qn} equally likely choices of the core maps. Averaging (6) and multiplying by this number proves (1). $\square$

Lemma 4. *For maps u, v of a finite set, $|\text{cyc}(u \circ v)| = |\text{cyc}(v \circ u)|$. For every positive integer d , $\text{cyc}(f^d) = \text{cyc}(f)$.*

Proof. The map v sends $\text{cyc}(u \circ v)$ to $\text{cyc}(v \circ u)$, by the intertwining identity $v(uv)^j = (vu)^j v$. Its restriction is injective: if $v(x) = v(y)$, then $uv(x) = uv(y)$, and uv is a permutation on its cyclic set. Interchanging u, v gives the reverse inequality of finite cardinalities. The assertion about powers follows directly: periodicity under f^d implies periodicity under f ; if $f^j(x) = x$, then $(f^d)^j(x) = x$. $\square$

Any word with a singleton letter may be cyclically rotated to VB . This lemma justifies the stated extension, without assuming that the distributions of the composite maps themselves are invariant under cyclic rotation.

3 A finite recursion on two trajectories

Let $k = |V|$, and let $L_1, \dots, L_k$ be the letters in the order in which they are applied, namely the reversal of V . Start with two distinguished vertices, numbered 0 and 1. A partial transition table has one row for each currently known vertex and one column for each distinct letter. Initially every entry is undefined. Reveal the length- k trajectory from 0 first, and then the length- k trajectory from 1, retaining the same table.

At a request for the transition of color L_j from vertex x :

1. If that table entry is already defined, follow it without branching or adding an exposed entry.
2. Otherwise, if there are v current vertices, branch into $v + 1$ possibilities. The target may be any of the v old vertices, or one new vertex canonically numbered v . In each branch record the transition and increase the number of exposed entries by one.

At termination record the number v of vertices and the number e of distinct exposed entries. Keep the leaf if the two final vertices coincide, and let $d_V(v, e)$ be the number of kept leaves with those statistics. This is a terminating integer recursion: at most $2k$ transitions are requested, and at most one new vertex is added at any request. In particular it is independent of n .

Proof of (2). For actual core maps and fixed distinct starting labels 1, 2, the two trajectories determine exactly one recursion leaf, by assigning canonical numbers to new labels in order of first appearance. Reusing a transition-table entry ensures that every repeated request uses the same map value.

Conversely, consider a leaf with v vertices and e exposed entries. Its $v - 2$ nondistinguished vertices can be assigned distinct labels from $[n] \setminus \{1, 2\}$ in exactly $(n - 2)_{v-2}$ ways. For each such assignment, the leaf prescribes e distinct input-letter entries of the random maps, so its probability is n^{-e} . These entries are mutually independent, even when some of their target labels coincide. The unexposed entries are unrestricted and sum to probability one. Different labelings or leaves describe disjoint events, and every pair of trajectories is represented. Summing the weights of exactly the leaves with coincident endpoints proves the formula. Since each falling factorial is an integer polynomial and there are finitely many terms, the Laurent-polynomial assertion follows. $\square$

An equivalent definition is useful for checking implementations. Take the $2k + 2$ formal positions of the two trajectories and partition them according to equality of their labels, keeping the two starting positions in distinct blocks. Retain only partitions for which any two transitions of the same color from the same block have targets in the same block. There is one recursion leaf for each retained partition. Its v is the number of blocks and its e is the number of distinct source-block/color pairs. This formulation also shows directly that $d_V(v, e)$ counts equality patterns, without automorphism or factorial factors.

4 Closed formulas for the first words

For $V = A$, the two values $a(1), a(2)$ are independent uniform labels, so $c_A(n) = 1/n$. Substitution in (1) proves (3).

For $V = AA$, write $x = a(1)$ and $y = a(2)$. If $x = y$, which has probability $1/n$, the two second iterates agree. Suppose $x \neq y$. Among the $n(n-1)$ ordered pairs (x, y) , the two pairs $(1, 2), (2, 1)$ give unequal second iterates. For every other pair, either both $a(x), a(y)$ are fresh independent uniform values, or exactly one is such a value and the other is fixed. Their conditional collision probability is $1/n$. Consequently

$$c_{AA}(n) = \frac{1}{n} + \frac{n(n-1)-2}{n^3} = \frac{2}{n} - \frac{1}{n^2} - \frac{2}{n^3}.$$

This gives (4), including $n = 2$.

As an additional value from the recursion,

$$\frac{T_{A^3B}(n)}{n^{2n}} = \frac{4n^5 - 4n^4 - 32n^3 + 123n^2 - 162n + 72}{n^6}.$$

The accompanying code also supplies the exact rational expressions for A^rB , $1 \leq r \leq 8$. These are evaluations of the proved recursion, not extrapolations from finitely many counts.

There is a second interpretation of (3). Place the domain of a and the domain of b in two labeled copies of $[n]$, using their values as arcs to the opposite copy. The AB -tree event says that the resulting bipartite functional digraph has just one cycle, of length two. Merging its two opposing arcs produces a spanning tree of $K_{n,n}$ with a distinguished edge. Conversely, orienting every edge towards the distinguished edge and replacing that edge by the two opposing arcs recovers the pair of maps. Every spanning tree has $2n-1$ edges, and the classical complete-bipartite tree count is n^{2n-2} . This gives the same formula. For $n \geq 2$, the latter tree count follows, for example, from the Laplacian eigenvalues $0, n$ (multiplicity $2n-2$), and $2n$, and the matrix-tree theorem.

5 Exact checks and scope

The source archive includes two different implementations. The first reveals transitions recursively and independently enumerates complete maps. It checks every pair of maps for $1 \leq n \leq 4$, $1 \leq r \leq 4$, including (6) for each fixed a , and every single map for the collision probabilities with $1 \leq n \leq 6$, $1 \leq r \leq 6$. For example:

n	T_{AB}	T_{A^2B}	T_{A^3B}	T_{A^4B}
2	12	12	12	12
3	405	477	477	477
4	28672	36352	38656	38656

It also checks a genuinely multiletter core ABA with a fresh letter C , by enumerating all triples for $n = 2, 3$. The tree counts are 56 and 15147, respectively.

The second implementation enumerates set partitions of all trajectory positions first and rejects inconsistent transitions. It recovers every diagram coefficient for the 15 binary application sequences of length at most four that begin with A , and checks normalization of all diagram weights for $2 \leq n \leq 10$. Direct map enumeration additionally checks 93 collision cases: all 31 binary application sequences beginning with A of length at most five, at $n = 2, 3, 4$. These finite checks support the implementation; the general identities follow from the proofs above.

We do not obtain a general formula for words without a singleton letter. Lemma 4 transfers the present counts to powers and rotations, but does not handle every word. The total-variation

questions discussed there and in [6] are not answered here. The neighboring permutation questions inadvertently bundled into UnsolvedMath record OWR-14299904-011 are also not addressed. Thus this note closes the explicitly proposed AB and AAB cases and gives a complete formula for the singleton-letter family; it does not close the whole dataset record.

The manuscript and exact-check programs were developed with AI assistance and checked by the authoring system. This preprint has not been independently refereed or formally verified. The underlying weighted-mapping and conditioning arguments are credited above; no absolute-priority claim is made.

References

- [1] G. Chapuy, Counting w -trees, in *Enumerative Combinatorics*, Oberwolfach Report 2/2026, pp. 145–146. doi:10.4171/OWR/2026/2.
- [2] G. Chapuy, *Short synchronizing words for random automata*, LPSM lecture slides, 2022, slide 48. <https://www.irif.fr/~chapuy/fichiers/slides/Chapuy-LPSM-2022.pdf>.
- [3] G. Chapuy and G. Perarnau, *Short Synchronizing Words for Random Automata*, arXiv:2207.14108v2 (2023). doi:10.1145/3736722.
- [4] A. Martinsson, *Synchronizing random automata through repeated ‘a’ inputs*, arXiv:2306.09040v2 (2023). <https://arxiv.org/abs/2306.09040>.
- [5] J. Pitman, Random mappings, forests, and subsets associated with Abel–Cayley–Hurwitz multinomial expansions, *Séminaire Lotharingien de Combinatoire* **46** (2001), Article B46h. <https://statistics.berkeley.edu/tech-reports/593>.
- [6] G. Chapuy and G. Perarnau, *Composition of random functions and word reconstruction*, arXiv:2603.28936v1 (2026). <https://arxiv.org/abs/2603.28936>.