The shortest word containing every permutation of ten symbols as a subsequence has length 83

Jason Davies*

November 18, 2026

hexagon:2611.00630v1

Abstract

A word is universal for an alphabet if it contains every permutation of the alphabet as a subsequence. We prove that no word of length 82 is universal for ten symbols. Together with the known construction of length 83, this shows that the shortest universal word on ten symbols has length 83. The proof cuts a word at the copies of one symbol: universality is equivalent to complementary prefix and suffix conditions at consecutive copies. A counting argument confines a least frequent symbol to four, five, six or seven copies and excludes four. In the remaining cases, frequency tests reduce the possibilities to finitely many block templates, and machine-checked propositional refutations exclude every ordering of each. The computation also reproduces the nine-symbol value 66.

AI transparency statement. The proof, its implementation and certificates, the Lean formalisation and this manuscript were produced by artificial intelligence under the direction of the human author. Details are given at the end of the paper.

1 The extremal problem

A word over an alphabet is *universal* if it contains every permutation of the alphabet as a subsequence, that is, at strictly increasing but not necessarily consecutive positions. Write $f(n)$ for the least length of a universal word over $[n]$. Requiring each permutation to occur contiguously gives the different superpermutation problem, whose answers grow factorially.

The problem is due to Karp [2, Problem 36]. Newey computed $f(n)$ through seven letters [8], and Adleman gave an early general construction [1]. Tan proved $f(8) = 52$ [12], and the OEIS credits Vitaliy Garnashevich with computing $f(9) = 66$ in 2025 [9, 6]. In general $f(n) = n^2 - o(n^2)$. Kleitman and Kwiatkowski proved $f(n) \geq n^2 - C_\varepsilon n^{7/4+\varepsilon}$ [7]. Radomirović proved $f(n) \leq \lceil n^2 - \frac{7}{3}n + \frac{19}{3} \rceil$ for $n \geq 7$ [10], and Tan obtained $n^2 - (\frac{5}{2} - \varepsilon)n + C_\varepsilon$ for infinitely many n [11]. Engen and Vatter survey the problem [5]. At ten letters Zălinescu records an earlier example of length 83 and gives a construction of length $n^2 - 2n + 3$ for all $n \geq 10$ [13].

Theorem 1. *The shortest universal word on ten symbols has length* 83.

Here is an upper-bound witness, written on two lines to be concatenated.

```
0123456789012345678029345167082345196078 21
3450697823140567821390456782310945678230 1
```

* jason@jasondavies.com

Matching each requested digit at its earliest available position verifies all 3,628,800 permutations. For the lower bound it suffices to exclude length 82, since appending letters preserves universality. The numerical inputs on smaller alphabets are

$$(f(0), \ldots, f(9)) = (0, 1, 3, 7, 12, 19, 28, 39, 52, 66).$$

The companion computation independently recovers the last entry from the preceding ones. Section 2 reduces the problem to four possible multiplicities of a least frequent symbol. Section 3 describes the tests that turn each case into a finite list of block templates, sequences of multisets whose concatenated orderings stand for candidate words. Section 4 treats the cases, summarised in Table 1.

Minimum frequency	Cases retained	Final check
4	Two outer cuts	Gap replacement gives a contradiction
5	922 block templates	922 ordering refutations
6	1,372 endpoint patterns	1,372 ordering refutations
7	1,131 endpoint templates	229 projected and 2 joint refutations

Table 1: The exclusion at length 82. The six-copy case also uses 304 smaller-alphabet frequency exclusions, proved with 777 auxiliary refutations.

2 A cut at consecutive copies

Write $\mathcal{U}(S, V)$ when V contains every ordering of the set S. Choose a symbol x with r occurrences in a word W, and let A be the alphabet with x removed. Delete all copies of x. In the resulting word, let P_i and Q_i be the portions before and after the location of the ith copy. Thus $|P_i| + |Q_i| = t = |W| - r$.

Lemma 2. *For $r \geq 1$, W is universal if and only if $\mathcal{U}(A, Q_1)$ and $\mathcal{U}(A, P_r)$ hold and, for every $1 \leq i < r$ and $S \subseteq A$,*

$$\mathcal{U}(S, P_i) \quad \text{or} \quad \mathcal{U}(A \setminus S, Q_{i+1}).$$

Proof. Requesting x first or last gives the outer conditions. If the disjunction fails, choose a missing ordering u of S and a missing ordering v of its complement. In uxv the pivot must come later than copy i and earlier than copy $i + 1$.

Conversely, for a given uxv, the copies of x that can follow u form a suffix of the occurrence list. Those that can precede v form a prefix. The two outer conditions make each list nonempty. If they are disjoint, the last copy in the prefix and the next copy contradict the disjunction for S, the alphabet of u. □

For example, with $A = \{a, b, c, d\}$, suppose $P_i = ab$ and $Q_{i+1} = dc$. The split $\{a, b\} \mid \{c, d\}$ fails because ba and cd are missing. The single request $baxcd$ rules out this cut regardless of the letters between the two pivot copies. We call a pair of words (P, Q) on A a *complementary cut* if $\mathcal{U}(S, P)$ or $\mathcal{U}(A \setminus S, Q)$ for every $S \subseteq A$. Thus each (P_i, Q_{i+1}) in a universal word is a complementary cut.

Lemma 3. *Suppose W is universal, $|A| = m$, and $F = f(m)$. Then*

$$F + m \;\leq\; r(t - F + m).$$

Proof. Write $h = t - F$. Each outer gap has length at most h, by the outer conditions. Across any internal gap, concatenate P_i, one arbitrary permutation of A, and Q_{i+1}. This word is universal on A. Indeed, embed the longest possible prefix of a requested ordering in P_i. If a letter remains,

take its next letter from the inserted permutation. The prefix through that letter was missing from P_i, so Lemma 2 puts every ordering of the complementary letters in Q_{i+1}. Replacing the gap therefore leaves at least F positions. Its original length was at most $h + m$. Summing gives $t \leq 2h + (r-1)(h+m)$, as required. □

At length 82 we obtain $r(25-r) \geq 75$ for every letter, which excludes $r \leq 3$. Choose a letter absent from the first nine positions. A permutation starting with that letter must read the other nine letters after those positions. At least 66 of the remaining 73 positions are therefore occupied by other letters. The minimum multiplicity lies in $\{4, 5, 6, 7\}$.

3 Tests on counts and on block orders

Capacity vectors. A capacity vector bounds the number of copies available for each letter of an alphabet. Insertion shows that realisability by a universal word is monotone in each coordinate. We use a recursive necessary condition C. On the empty alphabet it holds. On a nonempty alphabet B, it requires positive entries, a total of at least $f(|B|)$, and two further tests.

The *deletion test* uses the order of first appearances. Delete the prefix through the first appearance of a letter a. The remainder, with a erased, is universal on $B \setminus \{a\}$, and the deleted prefix used at least one copy of each letter appearing earlier. So charge one copy to each earlier letter and apply C recursively on $B \setminus \{a\}$. The test passes if, for some order of first appearances, every letter survives this deletion.

Lemma 4. *The condition C is monotone in each coordinate, and the deletion test passes for some order exactly when it passes for an order of decreasing capacity.*

Proof. Argue by induction on $|B|$. Interchange adjacent letters a, b with $c_a < c_b$ occurring in that order. The deletion of b improves when a is no longer charged. In the new deletion of a, the remaining capacity of b is at least the capacity of a in the old deletion of b. Other deletions have the same charges. By induction on the smaller alphabets, every deletion still passes. Monotonicity follows, since an admissible order remains available after a capacity increases. □

The *cut test* handles a pivot. On an alphabet of size m, tabulate a lower bound $H_m(p)$, capped at $f(m)$, on the length of Q in any complementary cut (P, Q) with $|P| = p$. A pair of capacity vectors with totals p, q passes if, for every S, C holds for S on the left or for the complement on the right. For a pivot with r copies and t other positions, iterate the upper bounds $u_1 = t - f(m)$ and $u_{i+1} = t - H_m(u_i)$ on $|P_i|$ while $u_i < f(m)$. The test fails if the last prefix has no room for $f(m)$ letters. The tables H_m are computed before testing a least frequent pivot on the next alphabet size, using the following alignment and discarding dominated subset signatures.

Lemma 5. *If a pair of vectors with given sorted entries passes the cut test in some alignment, it passes when both are aligned in increasing order.*

Proof. Fix the sorted left vector. Swap an inversion in the right vector on indices i, j with $c_i \leq c_j$ on the left. A right subset that could worsen contains i and omits j. If its complementary left test fails, replacing j there by i gives a weaker left vector, which also fails. The old right test on its complement thus supplies the needed right test after the swap. □

Cut catalogues. A cut catalogue lists the pairs of count vectors that can occur at a complementary cut with fixed budgets. It enumerates both vector partitions, applies the cut test in the sorted alignment, enumerates all relative alignments up to simultaneous relabelling, and retests the actual complementary subsets. When both budgets are below $f(8)$, both vectors on nine letters must be

positive, since a missing symbol on one side would require the other to be universal on eight letters. Padding a feasible cut on either side preserves feasibility.

Block templates. A block template is a sequence of multisets, the blocks. It stands for every word obtained by concatenating an ordering of each block. For a required alphabet inside a block interval, test its projected counts, then apply Lemma 2 at successive blocks containing each possible pivot. Each tested side includes its entire boundary block, which enlarges it. Recurse on the smaller alphabet. The test can also force an internal precedence $a < b$. If a requested prefix cannot finish by the preceding block containing a, and a disjoint requested suffix cannot start at or beyond the next block containing b, a request with those orders on either side of ab must read both letters in the current block. A missing letter or a cycle among singly occurring letters excludes the template.

Ordering formulas. For a surviving template, distinguish the individual occurrence tokens. Give differently labelled tokens in the same block Boolean order comparisons, fix each letter's copy order, and impose transitivity. Universality before a token on a nonempty set S requires, for each $a \in S$, an earlier a preceded by a prefix universal on $S \setminus \{a\}$. This is necessary because the latest available a can follow every prefix that any earlier a could follow. The reverse construction handles suffixes. Every universal ordering sets the comparisons, subset assertions, and witness variables consistently, so a refutation of the generated clauses rules out the template. For an exact two-sided cut the encoder can also start with short subset orders, test a proposed token order by greedy matching, and add the comparisons forced by a missing order. Each learned constraint excludes an actual failure of the cut condition.

4 Exclusion by minimum multiplicity

Lemma 6. *Every word universal on $k \geq 1$ letters has a letter of frequency at least k.*

Proof. For $k = 1$ this is immediate. For the induction step, delete the prefix through the first occurrence of the last new letter and erase that letter from the tail. The tail remains universal on the other $k - 1$ letters, and each of them lost at least one occurrence. □

Lemma 7. *Let (P, Q) be a complementary cut on nine letters.*

(a) If $|P| \leq 12$, then $|Q| \geq 50$.

(b) If $|P| \leq 11$, then $|Q| \geq 51$.

Proof. In both parts, if P misses a letter, Q must contain every order of the other eight, so $|Q| \geq 52$. Assume P contains all nine letters.

(a) At most three letters occur more than once in P. Choose six that occur once and call their set S. The other three occupy at most six positions of P. Since $f(3) = 7$, some order of those three is missing. Hence Q is universal on S and contains at least $f(6) = 28$ positions from S. By Lemma 6, the largest count u in Q on S is at least six. Let v be the second-largest. The two corresponding letters are singletons in P, so one of their orders is missing there. Thus Q also contains all permutations of their seven-letter complement. If $|Q| \leq 49$, then $u + v \leq 49 - f(7) = 10$ and $v \leq 4$. The six letters of S together occupy at most $u + v + 4v \leq 26$ positions, contradicting the required 28.

(b) Now P has at least seven singletons. For any six of them, the other three letters occupy at most five positions of P, so Q is universal on those six. By Lemma 6 applied to the six least frequent of the seven in Q, at least two of the seven have six or more copies in Q. They are singletons in

P, so one of their orders is missing there, and Q is universal on the other seven letters. Hence $|Q| \geq 6 + 6 + f(7) = 51$. □

Part (b) is sharp. Take $P = \texttt{80123456788}$ and

$$Q = \texttt{81756234083571642803657418235674081437256804713528 6}.$$

The word P is universal on 18 subsets. For each of the other 494 subsets, greedy matching verifies that Q is universal on its complement.

Four copies. Here $t = 78$ and $|P_1|, |Q_4| \leq 12$. Lemma 7(a) at the first cut gives $|Q_2| \geq 50$, hence $|P_2| \leq 28$. At the reverse end it gives $|P_3| \geq 50$, hence $|Q_3| \leq 28$. By the replacement argument in Lemma 3, P_2 followed by a permutation of the nine old letters and then Q_3 is universal on those letters. Its length is at most $28 + 9 + 28 = 65$, contradicting $f(9) = 66$.

Five copies. Now $t = 77$, so $|P_1|, |Q_5| \leq 11$. Lemma 7(b) gives $|Q_2|, |P_4| \geq 51$, hence $|P_2|, |Q_4| \leq 26$. Reverse if necessary so that $|Q_3| \leq 38$. At budgets $(26, 38)$, 287 central count pairs pass the frequency tests. For 22 of them the lazy cut search of Section 3 finds actual cut orders. The cut formulas of the other 265 have refutations. For any smaller pair, every one-letter padding towards the budgets must survive. Applying this condition successively leaves 22 pairs at $(26, 38)$, three at $(26, 37)$ and one at $(25, 38)$.

Write p, q, l for the count vectors of P_2, Q_3, P_3. Pad the reversed cut (Q_4, P_3) on its first side to 26. Since $|P_3| = 77 - |Q_3|$, the catalogues at $(26, 39)$ and $(26, 40)$ give the possibilities for l. Enumerate their relative labellings, retaining $g = l - p \geq 0$. At the outer cut, match the counts $g + q$ of Q_2 against the catalogues at $(11, 51)$ and $(11, 52)$, allowing shorter first sides by the padding rule. At $(11, 52)$ both vectors may be taken positive: if P_1 missed a letter, Q_2 would need all orders of the other eight and a copy of the missing letter, at least 53 positions, and a letter missing from Q_2 would likewise force $|P_1| \geq 52$. Write e for a surviving first-gap vector. The seven blocks are

$$e, \quad x, \quad p - e, \quad x, \quad g, \quad x, \quad q \cup \{x, x\},$$

where the last block consists of the letters counted by q together with the two remaining copies of x, in any order. Nonnegativity, the universal projected suffix at the first pivot, and the basic interval tests leave 1,669 templates. Each triple in the opening occupies at most five positions, so Q_2 must be universal on every complementary six-set. A pair of singletons in the opening likewise forces universality on its seven-letter complement. Apply the two-ended count test described below for six copies to the required interval projections, starting with an empty list of excluded profiles. This leaves 922 templates, each with an ordering refutation.

Six copies. The last new letter at either end occurs six or seven times. Before its first occurrence at most $82 - 6 - 66 = 10$ old letters can occur. Thus an opening through the last new letter has ten or eleven positions, with at most one repeated old letter. Enumerate both endpoint multisets. The last new letter alone, or two singleton letters in their missing order, force a universal complementary alphabet in the tail. The analogous right demands force prefix universality. A disjoint pair of left and right demands forces universality in the shared middle. Of 507 total profiles passing the deletion test, 299 give 3,082 endpoint patterns after these tests.

A stronger, two-ended count test uses a library of 304 excluded capacity vectors. For a vector on k letters, enumerate both last new letters and the counts used through their first occurrences. For a pivot of frequency r at total length L, the opening can contain at most $L - r - (k - 1) - f(k - 1)$ old letters beyond their mandatory first copies. Test the resulting tail, reversed tail and common

middle. Distinct last new letters give disjoint end blocks, since universality requires the left one to appear before the right one somewhere. A common last new letter occurring at least twice also gives disjoint blocks. For a single common occurrence the two tails must separately be universal on the other letters. The library is verified in increasing alphabet size. Each profile may use only already established exclusions on strictly smaller alphabets. Within an alphabet, excluding a vector rules out all componentwise smaller vectors by insertion.

k	6	7	8	9	total
excluded profiles	15	61	102	126	304
ordering refutations	66	360	329	22	777

This leaves 1,372 ten-letter endpoint patterns. Each describes five blocks, namely the opening before its last new letter, that letter, the unrestricted middle, the last new letter from the reverse end, and the closing remainder.

For these larger formulas the archive supplies trial words, from which the verifier regenerates clauses of three kinds.

- *Missing permutations.* A failed greedy embedding retains the comparisons that made each skipped occurrence too early. Keeping those comparisons keeps the requested permutation missing.
- *Failed cuts.* A failed complementary cut gives two infeasible count vectors. Enlarge each while it remains infeasible. A satisfying order must move some ranked copy across its pivot to exceed a resulting capacity.
- *Crowded regions.* Prescribe disjoint orders at the two ends of a required region. Their greedy endpoints bound the interval available for the complementary alphabet. If that interval has insufficient capacity, either an endpoint comparison or a comparison keeping enough copies outside the interval must change.

Each clause is necessary for universality. The encoder also tests the possible first and last appearances in required regions with small slack. Discarding redundant clauses only enlarges the set of assignments, so a refutation of the remaining conjunction suffices. All 1,372 conjunctions are refuted.

Seven copies. Let x be the last new letter. Both the nine letters before its first occurrence and the 66 old letters required afterwards attain their length bounds. Thus the opening ten positions form a permutation ending in x, and x occurs seven times. Delete the opening and all later copies of x. In the resulting universal word of length 66 on nine letters, every letter occurs at least six times. Its last new letter y occurs at most $66 - 8 - 52 = 6$ times. Equality gives another opening permutation on this projection. Reading from the other end yields corresponding roles z, u. Roles at opposite ends can coincide.

The deletion and eight-subset tests leave 26 total count profiles. Enumerate assignments of the distinguished letters, allowing coincidences across ends, and split the extra pivot copies among the endpoint blocks. The interval tests leave 1,131 templates. Each implies a nine-letter requirement at either end. After canonical relabelling there are 297 distinct projected requirements. Refuting 229 of them covers all but two templates. For those two, impose the requirements from both ends in the joint ten-letter order. When the prescribed demands are disjoint, every order of the remaining alphabet must fit between them. The two joint formulas are unsatisfiable. This concludes the proof of Theorem 1.

5 Verification and scope

The companion repository contains the enumerators, encoders, certificates, detailed derivations and replay instructions [3]. The replay checks the nine-letter witness and 118 nine-letter refutations. At ten letters, the five-, six- and seven-copy archives contain 3,567 refutations with 11,918,778 clause additions. These include the 777 smaller-alphabet cases and the 265 infeasible central five-copy cuts, whose missing-order clauses are validated before use. A count pair without a refutation is always retained, so no exclusion rests on a SAT solver's answer. Each addition is checked by assuming its negation and demanding a unit-propagation conflict before adjoining it, and each trace ends in the empty clause. Coverage is checked against regenerated case lists.

The Lean 4 [4] development proves the necessary direction of Lemma 2, the gap replacement, and the restriction of the minimum frequency to $\{4, 5, 6, 7\}$, with $f(9) \geq 66$ as an explicit hypothesis. It also proves Lemma 7 and the four-copy exclusion directly, with the lower bounds at alphabet sizes $2, 3, 6, 7, 8, 9$ as hypotheses. Thus the inference from a universal word of length 82 to a minimum frequency in $\{5, 6, 7\}$ is kernel-checked. As an independent cross-check of the four-copy case, an earlier search padded both central sides to 31 and found 88 count pairs. Lean also checks proofs of contradiction for the regenerated propositional cores of these 88 pairs and of the 265 central five-copy cuts. Table 2 summarises how each ingredient is established.

Ingredient	Established by
Witness of length 83	Direct check of all 3,628,800 permutations
$f(8) = 52$	Tan [12]; external input, not formalised
$f(9) = 66$	Companion replay: witness and 118 RUP refutations
Lemma 2 (necessary direction), Lemmas 3 and 7; four copies	Lean kernel, assuming the smaller values of f
Five, six and seven copies	Companion replay: enumerators and 3,567 RUP refutations

Table 2: How each ingredient of the proof is established.

Declaration of AI-generated research and authorship

This manuscript is the product of an AI-driven research process directed by the human author. GPT-6.2-Astra (OpenAI), used through Codex, developed the proof, including the mathematical arguments, the enumerators and encoders, the certificates and their verification, the Lean formalisation and the manuscript. Claude Opus 5.5 (Anthropic), used through Claude Code, reviewed and revised the manuscript, and implemented the certificates for the central five-copy cuts, the replay's progress reporting and the regression tests. The human author wrote no part of the manuscript.

The human author has checked the arguments and takes full responsibility for the content of this paper, including the correctness of its mathematics, and for the accuracy of this disclosure.

References

[1] Leonard Adleman. Short permutation strings. *Discrete Mathematics* 10(2):197–200, 1974. doi:10.1016/0012-365X(74)90116-2.

[2] Václav Chvátal, David A. Klarner and Donald E. Knuth. *Selected combinatorial research problems*. Stanford University Technical Report STAN-CS-72-292, 1972. http://infolab.stanford.edu/TR/CS-TR-72-292.html.

[3] Jason Davies. *Permutation supersequence length*: software, proof certificates and Lean source. https://github.com/jasondavies/permutation-supersequence-length. This paper describes release `v1.0`.

[4] Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28*, Lecture Notes in Computer Science, vol. 12699, pp. 625–635. Springer, 2021. doi:10.1007/978-3-030-79876-5_37.

[5] Michael Engen and Vincent Vatter. Containing all permutations. *American Mathematical Monthly* 128:4–24, 2021. doi:10.1080/00029890.2021.1835384.

[6] Vitaliy Garnashevich. *Solver for OEIS A136094*. https://github.com/vsg/A136094.

[7] Daniel J. Kleitman and David J. Kwiatkowski. A lower bound on the length of a sequence containing all permutations as subsequences. *Journal of Combinatorial Theory, Series A* 21(2):129–136, 1976. doi:10.1016/0097-3165(76)90057-1.

[8] Malcolm Newey. *Notes on a problem involving permutations as subsequences*. Stanford Artificial Intelligence Laboratory Memo AIM-190, STAN-CS-73-340, 1973. http://infolab.stanford.edu/TR/CS-TR-73-340.html.

[9] OEIS Foundation Inc. *The On-Line Encyclopedia of Integer Sequences*, sequence A062714. https://oeis.org/A062714.

[10] Saša Radomirović. A construction of short sequences containing all permutations of a set as subsequences. *Electronic Journal of Combinatorics* 19(4):#P31, 2012. doi:10.37236/2859.

[11] Oliver Tan. Skip letters for short supersequence of all permutations. *Discrete Mathematics* 345, Article 113070, 2022. doi:10.1016/j.disc.2022.113070.

[12] Oliver Tan. Conjecture on supersequence lower bound related to Connell sequence. arXiv:2501.11386 [math.CO], 2025. doi:10.48550/arXiv.2501.11386.

[13] Eugen Zălinescu. Shorter strings containing all k-element permutations. *Information Processing Letters* 111(12):605–608, 2011. doi:10.1016/j.ipl.2011.03.018.