Full Permutation Summaries

1 Mean language entropy compared to between-language permutation

Test	filename	meanPerm	p	z
All segments	AllLangs_allSegments.csv	0.66	< 0.0001	-77.61
Consonants	AllLangs_Consonants_allSegments.csv	0.70	< 0.0001	-68.96
Vowels	AllLangs_Vowels_allSegments.csv	0.47	< 0.0001	-40.98
Permute within families	AllLangs_allSegments_byFamily.csv	0.61	< 0.0001	-46.42
Permute within areas	AllLangs_allSegments_byArea.csv	0.64	< 0.0001	-62.56
Permute within families andareas	AllLangs_allSegments_byAreaAndFamily.csv	0.61	< 0.0001	-41.42
Unanalysable words	AllLangs_unanalyzable_allSegments.csv	0.73	< 0.0001	-17.10
Unanalysable words, permute within families	AllLangs_unanalyzable_allSegments_byFamily.csv	0.67	< 0.0001	-7.19
Unanalysable words, permute within areas	AllLangs_unanalyzable_allSegments_byArea.csv	0.67	< 0.0001	-6.96
Unanalysable words, permute within families and areas	AllLangs_unanalyzable_allSegments_byAreaAndFamily.csv	0.65	< 0.0001	-3.72

Table 1: Results for wh words, all segments From results folder ../Results/SimplifiedPhonology/PermutationResults/

Test	filename	meanPerm	p	z
All segments	AllLangs_firstSegments.csv	0.78	< 0.0001	-72.36
Consonants	AllLangs_Consonants_firstSegments.csv	0.82	< 0.0001	-61.92
Vowels	AllLangs_Vowels_firstSegments.csv	0.60	< 0.0001	-37.49
Permute within families	AllLangs_firstSegment_byFamily.csv	0.60	< 0.0001	-40.09
Permute within areas	AllLangs_firstSegment_byArea.csv	0.71	< 0.0001	-58.28
Permute within families andareas	AllLangs_firstSegment_byAreaAndFamily.csv	0.58	< 0.0001	-35.93
Unanalysable words	AllLangs_unanalyzable_firstSegments.csv	0.84	< 0.0001	-20.64
Unanalysable words, permute within families	AllLangs_unanalyzable_firstSegment_byFamily.csv	0.56	< 0.0001	-5.60
Unanalysable words, permute within areas	AllLangs_unanalyzable_firstSegment_byArea.csv	0.62	< 0.0001	-9.26
Unanalysable words, permute within families and areas	AllLangs_unanalyzable_firstSegment_byAreaAndFamily.csv	0.55	< 0.0001	-4.92

Table 2: Results for wh words, first segments From results folder ../Results/SimplifiedPhonology/PermutationResults/

Test	filename	meanPerm	p	z
All segments	AllLangs_allSegments_ActionDomain.csv	0.66	< 0.0001	-48.37
First segments	AllLangs_firstSegments_ActionDomain.csv	0.83	< 0.0001	-31.35
All segments, permute within families and areas	AllLangs_allSegments_BasicActionsDomain_byFamilyAndArea.csv	0.63	< 0.0001	-25.71
First segments, permute within families and areas	AllLangs_firstSegments_BasicActionsDomain_byFamilyAndArea.csv	0.75	< 0.0001	-15.30

Table 3: Results for basic action words From results folder ../Results/SimplifiedPhonology/PermutationResults/

Test	filename	meanPerm	p	z
All segments	AllLangs_allSegments_BodyDomain.csv	0.69	< 0.0001	-41.55
First segments	AllLangs_firstSegments_BodyDomain.csv	0.85	< 0.0001	-34.94
All segments, permute within families and areas	AllLangs_allSegments_BodyDomain_byFamilyAndArea.csv	0.66	< 0.0001	-21.15
First segments, permute within families and areas	AllLangs_firstSegments_BodyDomain_byFamilyAndArea.csv	0.74	< 0.0001	-13.64

Table 4: Results for body words From results folder ../Results/SimplifiedPhonology/PermutationResults/

Test	filename	meanPerm	p	z
All segments	AllLangs_allSegments_Pronouns.csv	0.72	< 0.0001	-35.90
First segments	AllLangs_firstSegments_Pronouns.csv	0.81	< 0.0001	-28.48
All segments, permute within families and areas	AllLangs_allSegments_PronounDomain_byFamilyAndArea.csv	0.67	< 0.0001	-21.12
First segments, permute within families and areas	AllLangs_firstSegments_PronounDomain_byFamilyAndArea.csv	0.71	< 0.0001	-15.71

Table 5: Results for pronouns From results folder ../Results/SimplifiedPhonology/PermutationResults/

Test	filename	meanPerm	p	z
All segments	Permutation_allSegments*	0.67	< 0.0001	-4.61
First segments	Permutation_firstSegments*	0.85	< 0.0001	-6.36
Same domain, permute within family, all segments	Permutation_Domain_byFamily_allSegments*	0.65	0.057	-1.49
Same domain, permute within family, first segments	Permutation_Domain_byFamily_firstSegments*	0.80	0.02	-2.52
Same domain, permute within family and area, all segments	Permutation_Domain_byFamily_and_Area_allSegments*	0.65	0.12	-1.18
Same domain, permute within family and area, first segments	Permutation_Domain_byFamily_and_Area_firstSegments*	0.79	0.041	-1.94

Table 6: Similarity of randomly selected concepts within a language, compared to between languages. From results folder ../Results/SimplifiedPhonology/PermutationResults/RandomConcepts/RandomConceptPermutationTest/

2 Compare entropy of wh words to other sets of words

Test	filename	meanPerm	p	z
All segments	Comparison_WH_Random_allSegments.csv	0.64	< 0.0001	-8.27
First segments	Comparison_WH_Random_firstSegments.csv	0.79	< 0.0001	-27.47
From same semantic domain, all segments	Comparison_WH_Domain_allSegments_concept*	0.63	< 0.0001	-5.64
From same semantic domain, first segments	Comparison_WH_Domain_firstSegments_concept*	0.76	< 0.0001	-16.13
Unanalysable words, all segments	Comparison_WH_Random_unanalyzable_allSegments.csv	0.71	< 0.0001	-4.96
Unanalysable words, first segments	Comparison_WH_Random_unanalyzable_firstSegments.csv	0.83	< 0.0001	-12.97
Initial languages only, all segments	Comparison_Initial_WH_Random_All_allSegments.csv	0.63	< 0.0001	-7.02
Initial languages only, first segments	Comparison_Initial_WH_Random_All_firstSegments.csv	0.78	< 0.0001	-22.38
Non-Initial languages only, all segments	Comparison_NonInitial_WH_Random_All_allSegments.csv	0.62	< 0.0001	-5.53
Non-Initial languages only, first segments	Comparison_NonInitial_WH_Random_All_firstSegments.csv	0.79	< 0.0001	-22.43

Table 7: Comparing the mean entropy of wh words to a randomly selected set of words.
From results folder ../Results/SimplifiedPhonology/PermutationResults/RandomConcepts/

3 Random independent samples tests

Test	filename	meanPerm	p	z
Unanalyzable wh words, permuting within families, all segments	RIS_WH_Unanalyzable_Family_allSegments.csv	0.10	0.97	-1.91
Unanalyzable wh words, permuting within families, first segments	RIS_WH_Unanalyzable_Family_firstSegments.csv	-0.24	0.00	2.84
Unanalyzable wh words, permuting within areas, all segments	RIS_WH_Unanalyzable_Area_allSegments.csv	0.09	0.96	-1.75
Unanalyzable wh words, permuting within areas, first segments	RIS_WH_Unanalyzable_Area_firstSegments.csv	-0.23	0.00	2.60
Body concepts, permuting within family, all segments	RIS_BodyConcepts_allSegments_Family.csv	-0.02	0.11	1.20
Body concepts, permuting within family, first segments	RIS_BodyConcepts_firstSegments_Family.csv	-0.14	0.00	2.48
Body concepts, permuting within area, all segments	RIS_BodyConcepts_allSegments_Area.csv	-0.01	0.37	0.32
Body concepts, permuting within area, first segments	RIS_BodyConcepts_firstSegments_Area.csv	-0.10	0.07	1.45
Action concepts, permuting within family, all segments	RIS_BasicActionsConcepts_allSegments_Family.csv	-0.01	0.39	0.27
Action concepts, permuting within family, first segments	RIS_BasicActionsConcepts_firstSegments_Family.csv	-0.03	0.27	0.62
Action concepts, permuting within area, all segments	RIS_BasicActionsConcepts_allSegments_Area.csv	-0.00	0.44	0.14
Action concepts, permuting within area, first segments	RIS_BasicActionsConcepts_firstSegments_Area.csv	-0.01	0.43	0.22
Pronouns, permuting within family, all segments	RIS_PronounConcepts_allSegments_Family.csv	-0.04	0.04	1.69
Pronouns concepts, permuting within family, first segments	RIS_PronounConcepts_firstSegments_Family.csv	-0.01	0.41	0.22
Pronouns concepts, permuting within area, all segments	RIS_PronounConcepts_allSegments_Area.csv	-0.03	0.18	0.94
Pronouns concepts, permuting within area, first segments	RIS_PronounConcepts_firstSegments_Area.csv	-0.03	0.33	0.46
Random concepts, all segments	RIS_RandomConcepts_allSegments*	-0.01	0.37	0.33
Random concepts, first segments	RIS_RandomConcepts_firstSegments*	-0.01	0.40	0.27
Random concepts within the same domain, all segments	RIS_RandomConcepts_Domain_allSegments*	-0.00	0.41	0.17
Random concepts within the same domain, first segments	RIS_RandomConcepts_Domain_firstSegments*	-0.03	0.27	0.61

Table 8: Random independent samples tests, comparing initial interrogative languages and non-initial interrogative languages. From results folder ../Results/SimplifiedPhonology/PermutationResults/RandomIndependentSamples/

Test	filename	meanPerm	p	z
Wh words, all consonants	ConsonantsInitial_3_allSegments_RandomIndependentSample.csv	-0.03	0.07	1.48
Wh words, first consonant	ConsonantsInitial_3_firstSegments_RandomIndependentSample.csv	-0.12	0.03	1.89
Wh words, all vowels	VowelsInitial_3_allSegments_RandomIndependentSample.csv	0.01	0.61	-0.29
Wh words, first vowel	VowelsInitial_3_firstSegments_RandomIndependentSample.csv	-0.08	0.04	1.80

Table 9: Random independent samples tests, comparing wh words by consonants or vowels separately (controlling for language family). From results folder ../Results/SimplifiedPhonology/PermutationResults/RandomIndependentSamples/

Test	filename	meanPerm	p	z
Wh words, all consonants	ConsonantsInitial_3_allSegments_Area_RandomIndependentSample.csv	-0.00	0.43	0.18
Wh words, first consonant	ConsonantsInitial_3_firstSegments_Area_RandomIndependentSample.csv	-0.05	0.26	0.65
Wh words, all vowels	VowelsInitial_3_allSegments_Area_RandomIndependentSample.csv	0.03	0.81	-0.87
Wh words, first vowel	VowelsInitial_3_firstSegments_Area_RandomIndependentSample.csv	-0.05	0.23	0.74

Table 10: Random independent samples tests, comparing wh words by consonants or vowels separately (controlling for geogrpahic area). From results folder ../Results/SimplifiedPhonology/PermutationResults/RandomIndependentSamples/

4 Tests without duplicates

In several languages the same form is listed under more than one question-word concept. This decreases the entropy score, so it is reasonable to ask whether the results are driven by this effect. However, this is a difficult criticism to address. In an extreme reading, our hypothesis would predict that a language would have identical forms for all wh-words. Therefore, removing the duplicates removes part of the effect we are trying to detect.

Also, we made the following assumptions about the data: Empty cells indicate that the language has no lexicalised form for the concept. Duplicated forms mean that they use the same form for both concepts. So, if the language really only had one concept for *how* and *what*, then it would receive only one entry. Duplicate entries suggest that speakers have separate concepts, but identical forms. In this case, we think it is fair to count them as separate entries.

Actually implementing this check is also difficult. A lot of time was put into cleaning and simplifying the representations of words, often on a language-by-language basis. Forms that look identical in the final data may actually be phonemically different in their raw form. It is also not clear, if a language has a duplicate form, which concept to exclude (making study 1 difficult to repeat exactly).

Still, the results below are based on a dataset with duplicates within languages removed (conservatively based on the cleaned forms, see testDuplicated.R). Duplicate forms for other concepts were not removed.

When removing duplicates, the mean entropy of wh words first segments was 0.48 (compared to 0.46 with duplicates), and when looking at all segments it was 0.59 (compared to 0.57 with duplicates). The results below do not differ much from the original results, so we conclude that the effects in the main paper are not driven by an artefact of duplicated forms.

Test	filename	meanPerm	p	z
Wh vs Random Concepts				
...All segments	RandomConcepts/Comparison_WH_Random_allSegments_noDuplicates.csv	0.64	< 0.0001	-5.78
...First segmentsAll segments	RandomConcepts/Comparison_WH_Random_firstSegments_noDuplicates.csv	0.79	< 0.0001	-24.76
Random Independet Samples				
..By family				
...All segments	RandomIndependentSamples/InterrogativeOrder_RandomIndependentSamples_allSegments_noDuplicates.csv	-0.01	0.27	0.6
...First segments	RandomIndependentSamples/InterrogativeOrder_RandomIndependentSamples_firstSegments_noDuplicates.csv	-0.074	0.12	1.19
..By area				
...All segments	RandomIndependentSamples/InterrogativeOrder_RandomIndependentSamples_allSegments_Areas_noDuplicates.csv	0.01	0.68	-0.47
...First segments	RandomIndependentSamples/InterrogativeOrder_RandomIndependentSamples_firstSegments_Areas_noDuplicates.csv	-0.054	0.26	0.65

Table 11: Results with duplicated forms removed within languages From results folder ../Results/SimplifiedPhonology/PermutationResults/

5 Controlling for number of words in permutation tests

The permutation tests where interrogative words were compared to random words worked by keeping the number of concepts equal in both sets. So, for example, the E_f score would be calculated for the entries for 9 interrogative concepts and for 9 random concepts. One concern here is that this might compare different numbers of words. For example, the 9 interrogative concepts might contain 10 unique words (due to multiple words in 1 entry), while the 9 random concepts might contain 8 unique words (due to missing data). Since larger samples are more likely to produce extreme values, it's possible that this was affecting the results.

(We note that often the interrogative set had more words per language than many of the other sets, and in this case a random additional word is more likely to increase the entropy than decrease it, so the prediction would be that controlling for number of concepts is conservative)

We re-ran the test from study 2, but keeping the number of unique words in each language the same. So, for example, if language X had 8 unique interrogative words, then 8 words were selected from the random set. However, the languages in the sample had between 5 and 25 unique words for interrogatives (median = 10.8, sd = 3), so it was impossible to keep both the number of concepts and the number of words equal. So, the procedure worked like this: For each language L_i from L_1 to L_{226}, the number of unique interrogative words was obtained, call this U_{L_i}. The data for random concepts was organised as a matrix, with rows representing concepts and columns representing languages. The order of rows was randomly permuted. For each column, the first U_{L_i} words in the matrix were taken as the corresponding set of words. In this way, each language had a set of interrogative words, and an equal number of random words, with most languages being matched on concepts as far as it was possible.

Table 5 below shows the results of the tests, comparing the original tests (controlling for number of concepts) to the test controlling for number of words. The latter results in a more extreme result than the original (more in favour of our hypothesis), for both first segments and all segments. Thus, we argue that the results in the paper are not driven by comparing different numbers of words.

	Controlling for ...	p	z
First segments	Number of concepts	< 0.001	-27.93
	Number of words	< 0.001	-34.13
All segments	Number of concepts	< 0.001	-8.49
	Number of words	< 0.001	-9.34

We also performed an additional test comparing the mean E_f for interrogative words to the mean E_f for pronouns. In this test, a sample of words from each set was taken for each language, so that the size of the sample was the minimum number of words available. So, if a language had 9 interrogative words and 6 pronoun words, a random sample of 6 interrogative words was taken from the full set of 9. Mean E_f was calculated for both sets, and the difference taken. This was repeated 1000 times.

As in the original experiment, we found that interrogative words were more similar (smaller E_f) than pronouns (for first segments: mean difference in E_f = -0.15, $p < 0.001$, $z = 24.67$).