# Full Permutation Summaries

## 1 Mean language entropy compared to between-language permutation

| Test | filename | meanPerm | p | z |
|---|---|---|---|---|
| All segments | AllLangs_allSegments.csv | 0.65 | < 0.0001 | -59.02 |
| Consonants | AllLangs_Consonants_allSegments.csv | 0.69 | < 0.0001 | -53.28 |
| Vowels | AllLangs_Vowels_allSegments.csv | 0.47 | < 0.0001 | -27.49 |
| Permute within families | AllLangs_allSegments_byFamily.csv | 0.61 | < 0.0001 | -28.84 |
| Permute within areas | AllLangs_allSegments_byArea.csv | 0.64 | < 0.0001 | -47.14 |
| Permute within families andareas | AllLangs_allSegments_byAreaAndFamily.csv | 0.60 | < 0.0001 | -24.07 |
| Unanalysable words | AllLangs_unanalyzable_allSegments.csv | 0.75 | < 0.0001 | -10.65 |
| Unanalysable words, permute within families | AllLangs_unanalyzable_allSegments_byFamily.csv | 0.70 | < 0.0001 | -4.99 |
| Unanalysable words, permute within areas | AllLangs_unanalyzable_allSegments_byArea.csv | 0.70 | < 0.0001 | -4.86 |
| Unanalysable words, permute within families and areas | AllLangs_unanalyzable_allSegments_byAreaAndFamily.csv | 0.69 | 0.0004 | -3.39 |

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 | -64.49 |
| Consonants | AllLangs_Consonants_firstSegments.csv | 0.80 | < 0.0001 | -55.08 |
| Vowels | AllLangs_Vowels_firstSegments.csv | 0.60 | < 0.0001 | -30.36 |
| Permute within families | AllLangs_firstSegment_byFamily.csv | 0.57 | < 0.0001 | -31.29 |
| Permute within areas | AllLangs_firstSegment_byArea.csv | 0.71 | < 0.0001 | -52.93 |
| Permute within families andareas | AllLangs_firstSegment_byAreaAndFamily.csv | 0.54 | < 0.0001 | -26.65 |
| Unanalysable words | AllLangs_unanalyzable_firstSegments.csv | 0.85 | < 0.0001 | -15.90 |
| Unanalysable words, permute within families | AllLangs_unanalyzable_firstSegment_byFamily.csv | 0.61 | < 0.0001 | -4.92 |
| Unanalysable words, permute within areas | AllLangs_unanalyzable_firstSegment_byArea.csv | 0.66 | < 0.0001 | -7.37 |
| Unanalysable words, permute within families and areas | AllLangs_unanalyzable_firstSegment_byAreaAndFamily.csv | 0.60 | < 0.0001 | -5.01 |

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 | 0.65 | < 0.0001 | -37.81 |
| First segments | AllLangs_firstSegments_ActionDomain.csv | 0.83 | < 0.0001 | -25.63 |
| All segments, permute within families and areas | AllLangs_allSegments_BasicActionsDomain_byFamilyAndArea.csv | 0.62 | < 0.0001 | -14.54 |
| First segments, permute within families and areas | AllLangs_firstSegments_BasicActionsDomain_byFamilyAndArea.csv | 0.75 | < 0.0001 | -11.28 |

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 | -37.11 |
| First segments | AllLangs_firstSegments_BodyDomain.csv | 0.85 | < 0.0001 | -33.08 |
| All segments, permute within families and areas | AllLangs_allSegments_BodyDomain_byFamilyAndArea.csv | 0.65 | < 0.0001 | -17.08 |
| First segments, permute within families and areas | AllLangs_firstSegments_BodyDomain_byFamilyAndArea.csv | 0.72 | < 0.0001 | -12.39 |

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

| Test | filename | meanPerm | p | z |
|---|---|---|---|---|
| All segments | AllLangs_allSegments_Pronouns.csv | 0.70 | < 0.0001 | -31.72 |
| First segments | AllLangs_firstSegments_Pronouns.csv | 0.81 | < 0.0001 | -24.91 |
| All segments, permute within families and areas | AllLangs_allSegments_PronounDomain_byFamilyAndArea.csv | 0.65 | < 0.0001 | -15.61 |
| First segments, permute within families and areas | AllLangs_allSegments_PronounDomain_byFamilyAndArea.csv | 0.65 | < 0.0001 | -15.61 |

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

| Test | filename | meanPerm | p | z |
|---|---|---|---|---|
| All segments | Permutation_allSegments* | 0.67 | < 0.0001 | -4.69 |
| First segments | Permutation_firstSegments* | 0.84 | < 0.0001 | -7.42 |
| Same domain, permute within family, all segments | Permutation_Domain_byFamily_allSegments* | 0.64 | 0.056 | -1.39 |
| Same domain, permute within family, first segments | Permutation_Domain_byFamily_firstSegments* | 0.79 | 0.015 | -2.52 |
| Same domain, permute within family and area, all segments | Permutation_Domain_byFamily_and_Area_allSegments* | 0.64 | 0.14 | -1.06 |
| Same domain, permute within family and area, first segments | Permutation_Domain_byFamily_and_Area_firstSegments* | 0.78 | 0.043 | -1.88 |

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 | 0.63 | $< 0.0001$ | -7.03 |
| First segments | Comparison_WH_Random_firstSegments.csv | 0.78 | $< 0.0001$ | -28.10 |
| From same semantic domain, all segments | Comparison_WH_Domain_allSegments_concept* | 0.62 | $< 0.0001$ | -4.29 |
| From same semantic domain, first segments | Comparison_WH_Domain_firstSegments_concept* | 0.75 | $< 0.0001$ | -17.97 |
| Unanalysable words, all segments | Comparison_WH_Random_unanalyzable_allSegments.csv | 0.72 | 0.0037 | -2.60 |
| Unanalysable words, first segments | Comparison_WH_Random_unanalyzable_firstSegments.csv | 0.84 | $< 0.0001$ | -10.43 |
| Initial languages only, all segments | Comparison_Initial_WH_Random_All_allSegments.csv | 0.63 | $< 0.0001$ | -6.74 |
| Initial languages only, first segments | Comparison_Initial_WH_Random_All_firstSegments.csv | 0.78 | $< 0.0001$ | -22.21 |
| Non-Initial languages only, all segments | Comparison_NonInitial_WH_Random_All_allSegments.csv | 0.62 | $< 0.0001$ | -4.49 |
| Non-Initial languages only, first segments | Comparison_NonInitial_WH_Random_All_firstSegments.csv | 0.79 | $< 0.0001$ | -20.32 |

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 | 0.04 | 0.83 | -0.94 |
| Unanalyzable wh words, permuting within families, first segments | RIS_WH_Unanalyzable_Family_firstSegments.csv | -0.28 | 0.00 | 2.93 |
| Unanalyzable wh words, permuting within areas, all segments | RIS_WH_Unanalyzable_Area_allSegments.csv | 0.04 | 0.82 | -0.93 |
| Unanalyzable wh words, permuting within areas, first segments | RIS_WH_Unanalyzable_Area_firstSegments.csv | -0.28 | 0.00 | 2.83 |
| Body concepts, permuting within family, all segments | RIS_BodyConcepts_allSegments_Family.csv | -0.01 | 0.16 | 0.99 |
| Body concepts, permuting within family, first segments | RIS_BodyConcepts_firstSegments_Family.csv | -0.12 | 0.01 | 2.22 |
| Body concepts, permuting within area, all segments | RIS_BodyConcepts_allSegments_Area.csv | -0.01 | 0.38 | 0.30 |
| Body concepts, permuting within area, first segments | RIS_BodyConcepts_firstSegments_Area.csv | -0.09 | 0.10 | 1.28 |
| Action concepts, permuting within family, all segments | RIS_BasicActionsConcepts_allSegments_Family.csv | -0.01 | 0.25 | 0.70 |
| Action concepts, permuting within family, first segments | RIS_BasicActionsConcepts_firstSegments_Family.csv | -0.03 | 0.30 | 0.55 |
| Action concepts, permuting within area, all segments | RIS_BasicActionsConcepts_allSegments_Area.csv | -0.00 | 0.48 | 0.06 |
| Action concepts, permuting within area, first segments | RIS_BasicActionsConcepts_firstSegments_Area.csv | -0.03 | 0.36 | 0.39 |
| Pronouns, permuting within family, all segments | RIS_PronounConcepts_allSegments_Family.csv | -0.03 | 0.12 | 1.16 |
| Pronouns concepts, permuting within family, first segments | RIS_PronounConcepts_firstSegments_Family.csv | 0.02 | 0.69 | -0.48 |
| Pronouns concepts, permuting within area, all segments | RIS_PronounConcepts_allSegments_Area.csv | -0.03 | 0.17 | 0.95 |
| Pronouns concepts, permuting within area, first segments | RIS_PronounConcepts_firstSegments_Area.csv | -0.02 | 0.37 | 0.33 |
| Random concepts, all segments | RIS_RandomConcepts_allSegments* | -0.00 | 0.48 | 0.07 |
| Random concepts, first segments | RIS_RandomConcepts_firstSegments* | -0.01 | 0.39 | 0.27 |
| Random concepts within the same domain, all segments | RIS_RandomConcepts_Domain_allSegments* | 0.00 | 0.56 | -0.15 |
| Random concepts within the same domain, first segments | RIS_RandomConcepts_Domain_firstSegments* | -0.02 | 0.35 | 0.39 |

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 | -0.15 | 0.01 | 2.37 |
| Wh words, all vowels | VowelsInitial_3_allSegments_RandomIndependentSample.csv | -0.02 | 0.21 | 0.82 |
| Wh words, first vowel | VowelsInitial_3_firstSegments_RandomIndependentSample.csv | -0.07 | 0.05 | 1.67 |

Table 9: Random independent samples tests, comparing wh words by consonants or vowels separately. 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?s 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.

The mean entropy of wh words first segments was 0.46 (compared to 0.43 with duplicates), and when looking at all segments it was 0.59 (compared to 0.58 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 | 0.63 | $< 0.0001$ | -5.18 |
| ...First segmentsAll segments | RandomConcepts/Comparison_WH_Random_firstSegments_noDuplicates.csv | 0.78 | $< 0.0001$ | -26.23 |
| Random Independet Samples | | | | |
| ..By family | | | | |
| ...All segments | RandomIndependentSamples/InterrogativeOrder_RandomIndependentSamples_allSegments_noDuplicates.csv | -0.018 | 0.17 | 0.97 |
| ...First segments | RandomIndependentSamples/InterrogativeOrder_RandomIndependentSamples_firstSegments_noDuplicates.csv | -0.088 | 0.072 | 1.46 |
| ..By area | | | | |
| ...All segments | RandomIndependentSamples/InterrogativeOrder_RandomIndependentSamples_allSegments_Areas_noDuplicates.csv | -0.00022 | 0.49 | 0.01 |
| ...First segments | RandomIndependentSamples/InterrogativeOrder_RandomIndependentSamples_firstSegments_Areas_noDuplicates.csv | -0.051 | 0.24 | 0.68 |

Table 10: Results with duplicated forms removed within languages From results folder ../Results/SimplifiedPhonology/PermutationResults/