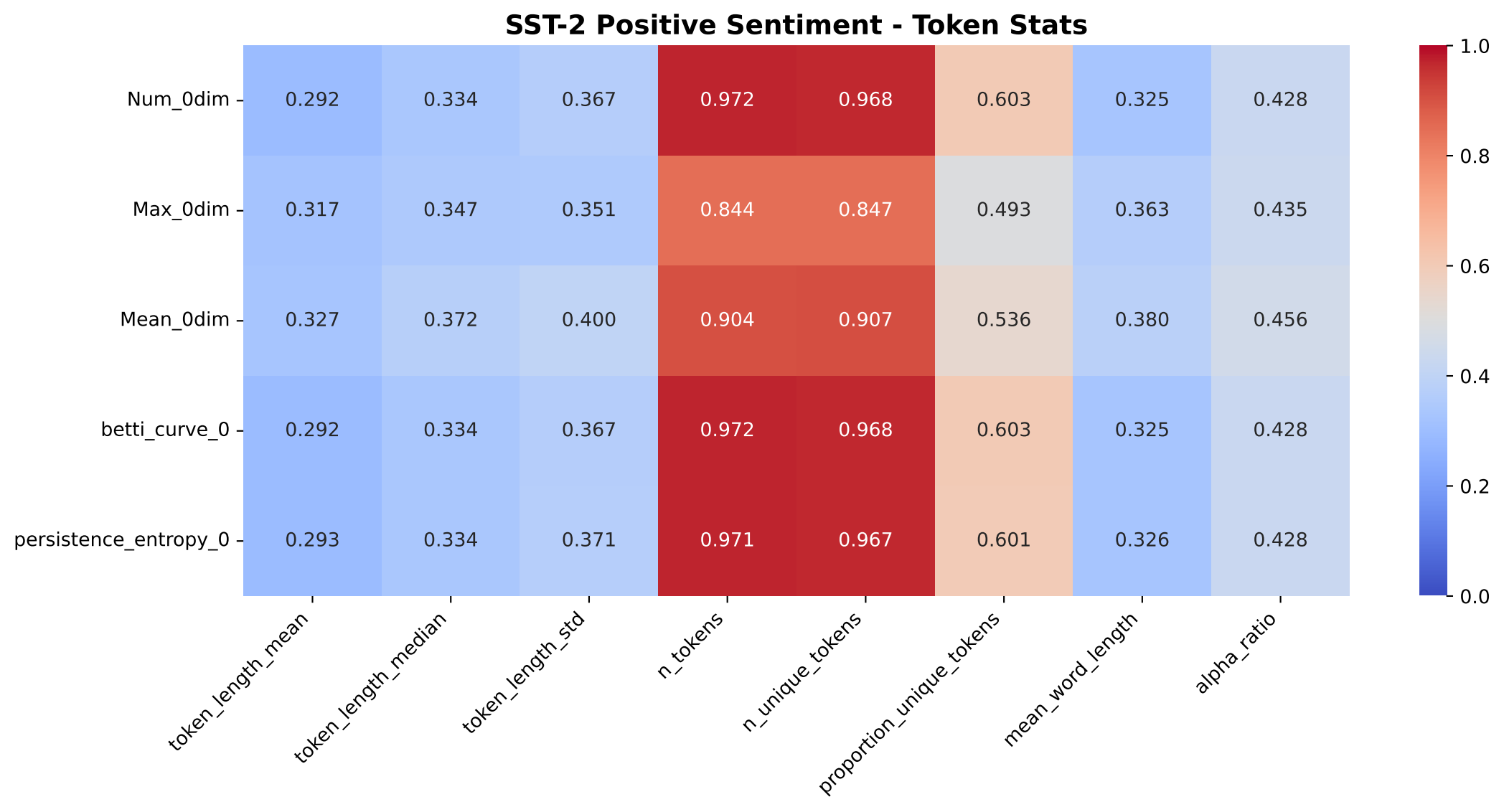
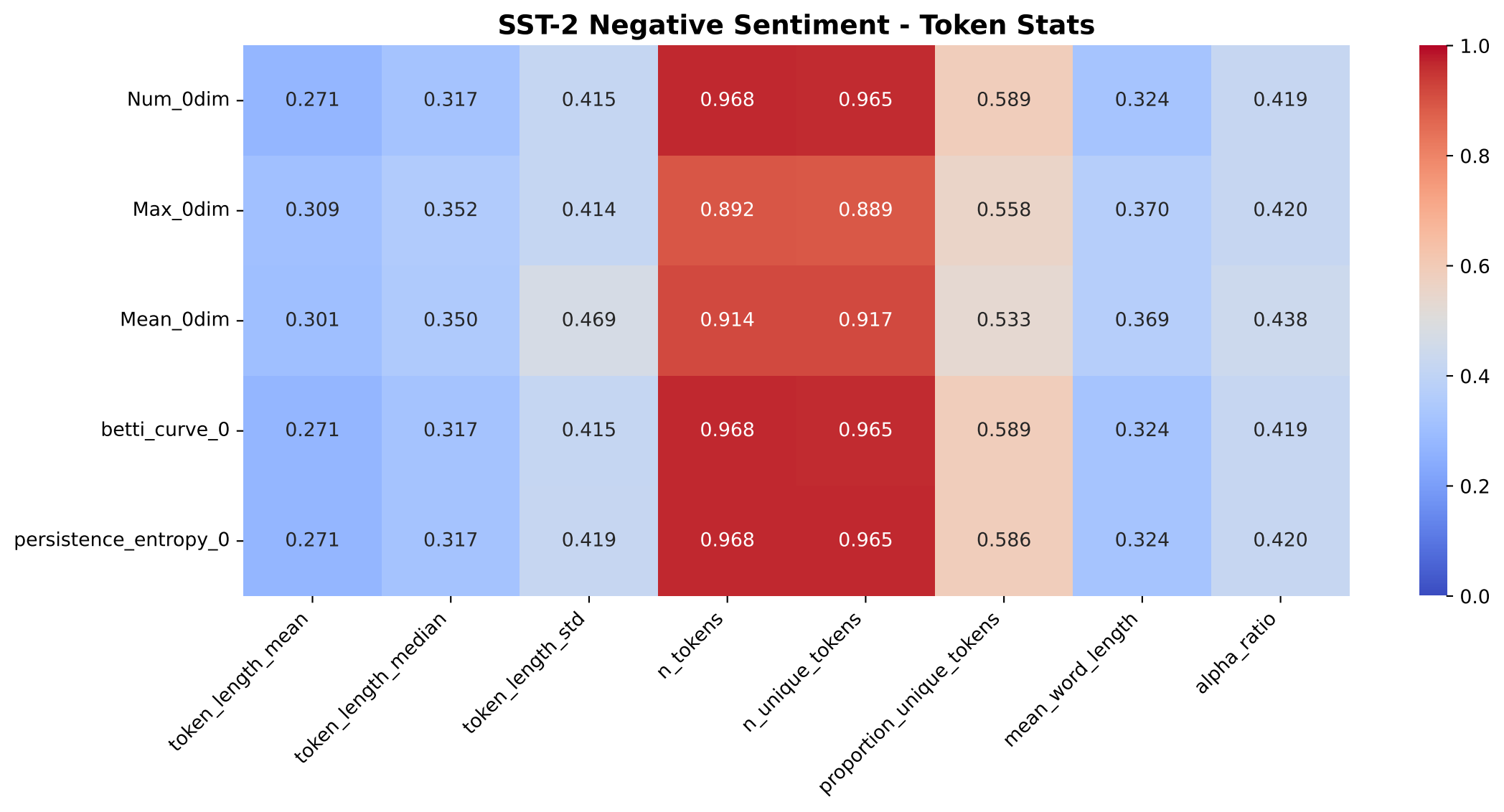
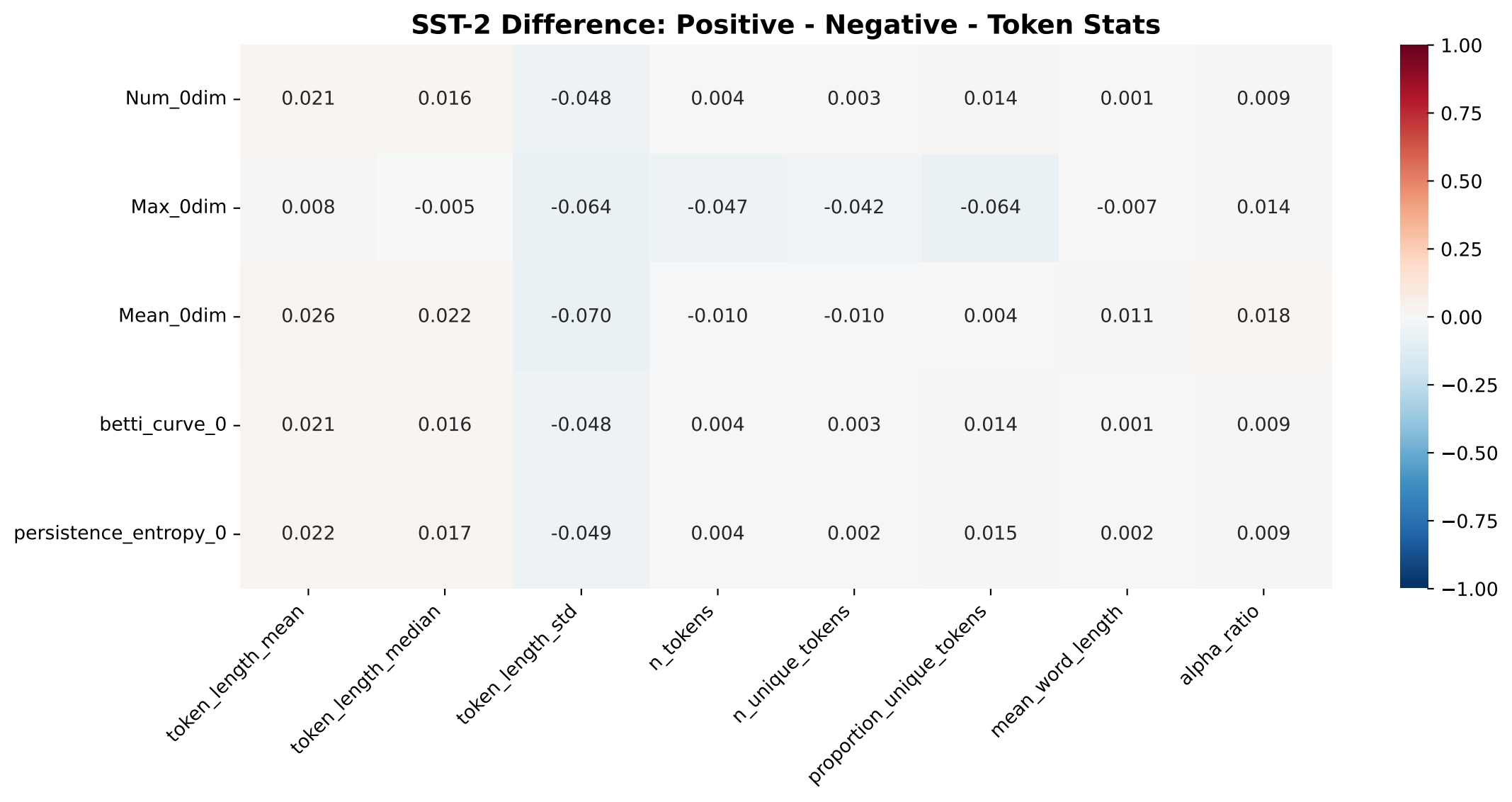
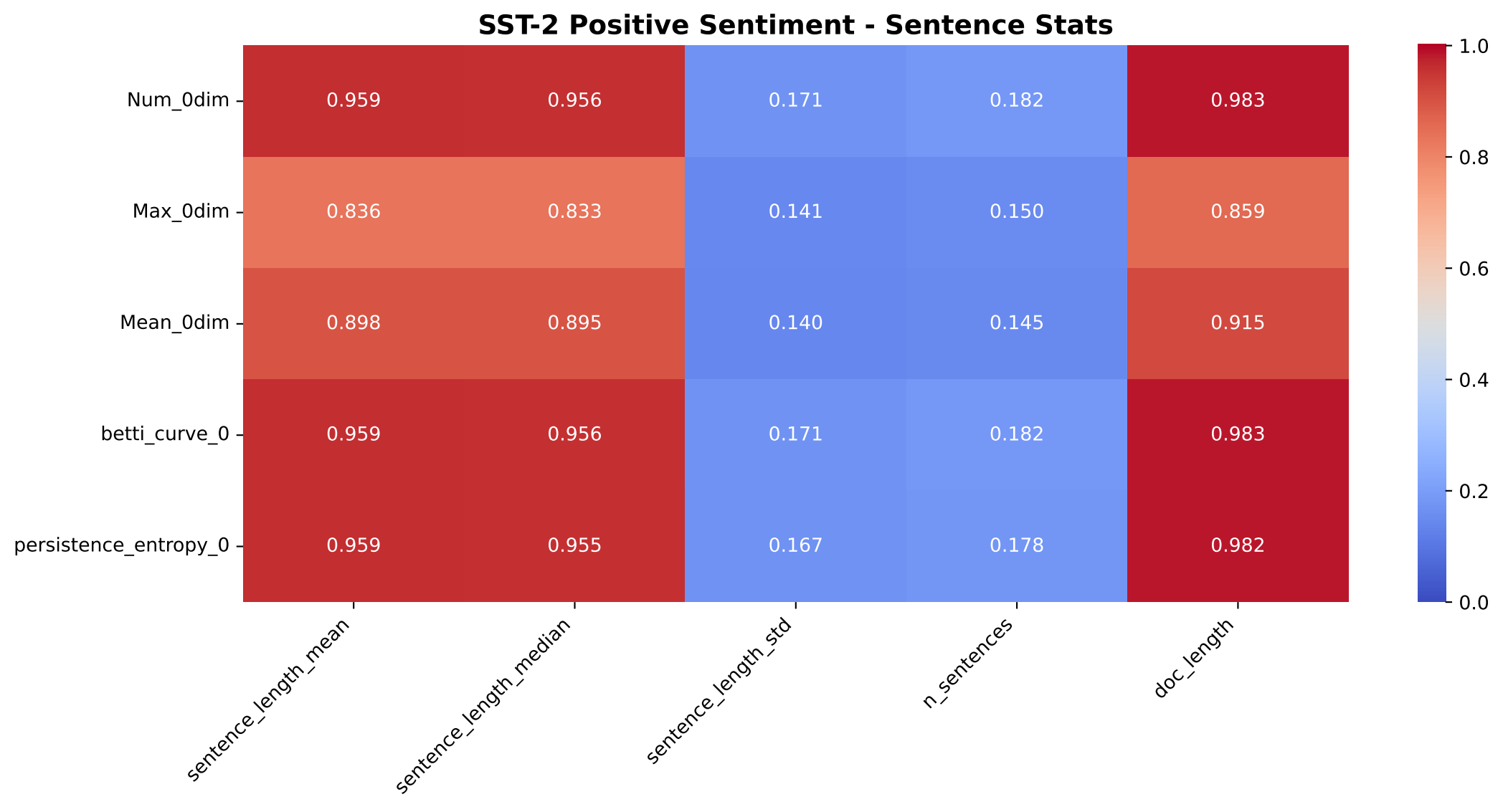
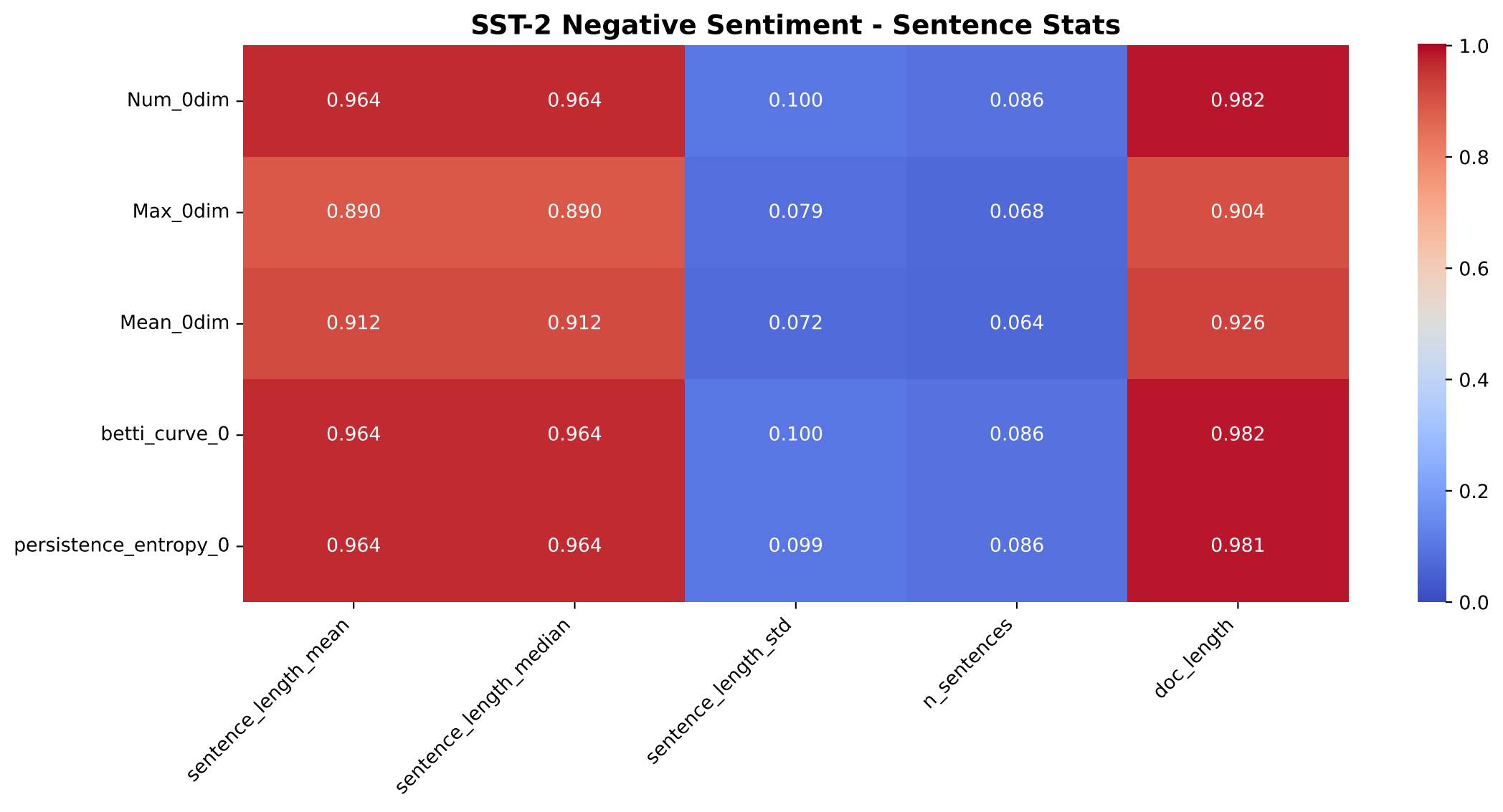
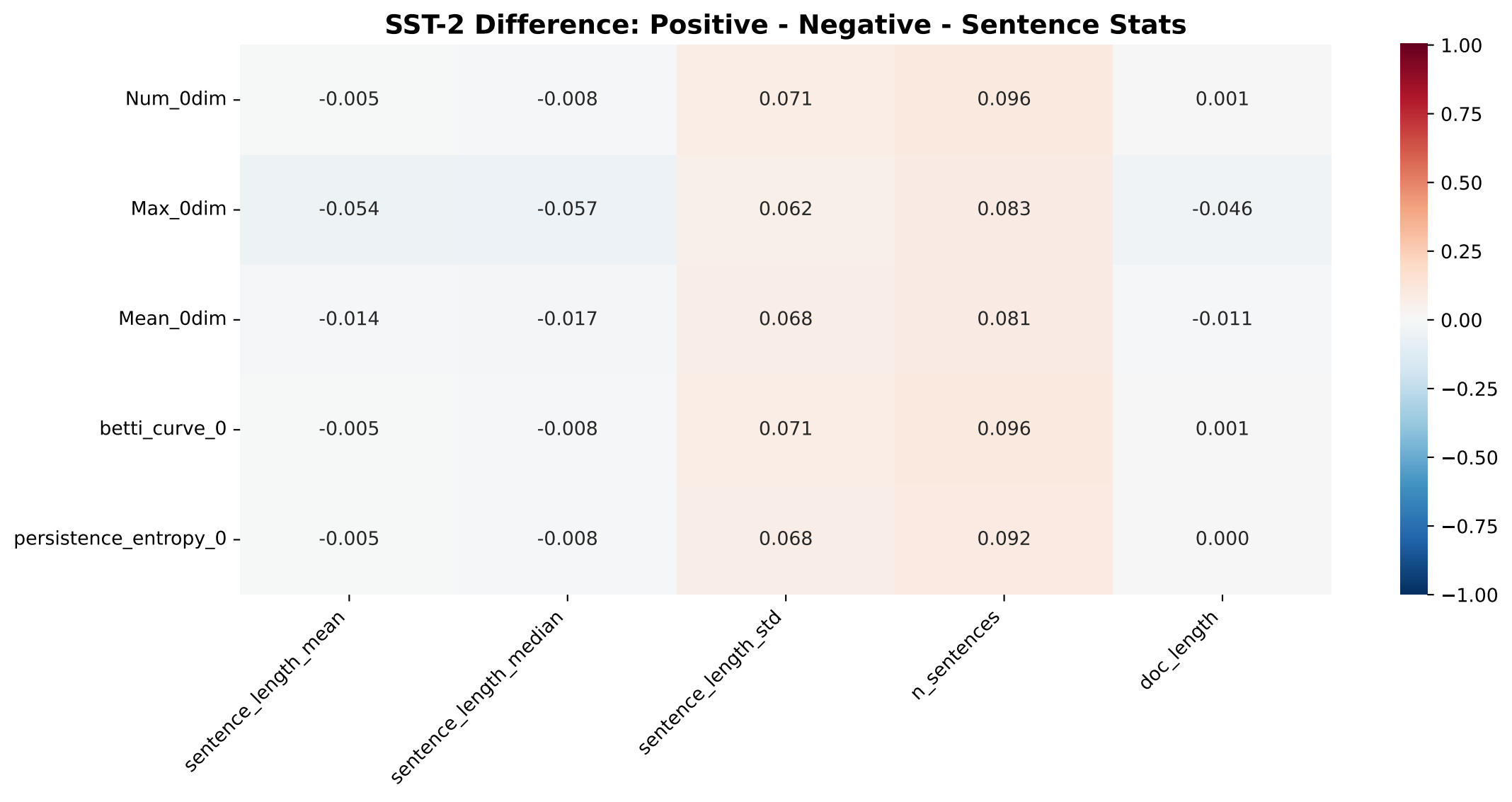

SST-2 Positive Sentiment - Token Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.292 0.334 0.367 0.972 0.968 0.603 0.325 0.428
0.317 0.347 0.351 0.844 0.847 0.493 0.363 0.435
0.327 0.372 0.400 0.904 0.907 0.536 0.380 0.456
0.292 0.334 0.367 0.972 0.968 0.603 0.325 0.428
0.293 0.334 0.371 0.971 0.967 0.601 0.326 0.428
token_length_mean
token_length_median
token_length_std
n_tokens
n_unique_tokens
proportion_unique_tokens
mean_word_length
alpha_ratio
1.0
0.8
0.6
0.4
0.2
0.0

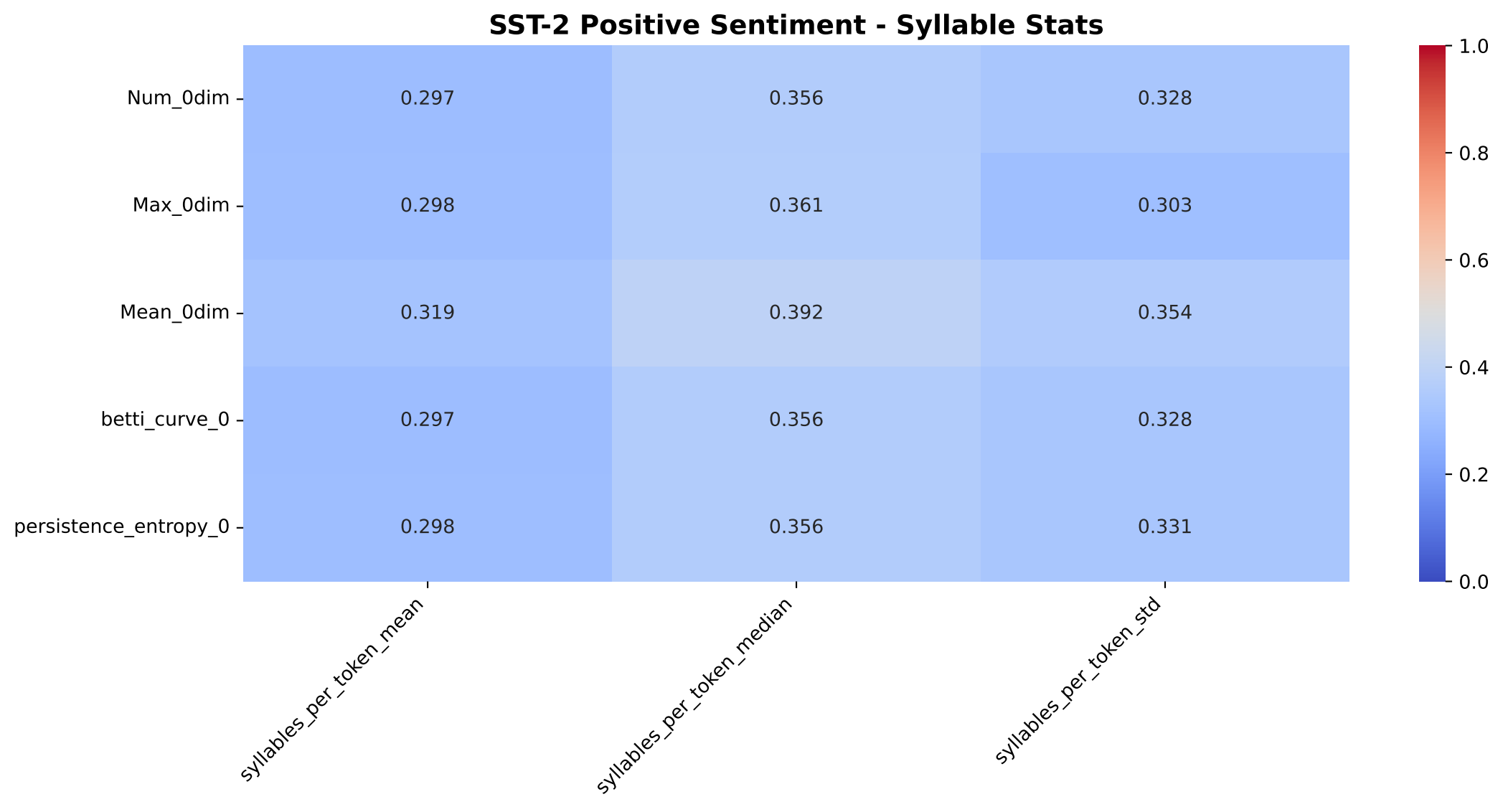

SST-2 Negative Sentiment - Token Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.271 0.317 0.415 0.968 0.965 0.589 0.324 0.419
0.309 0.352 0.414 0.892 0.889 0.558 0.370 0.420
0.301 0.350 0.469 0.914 0.917 0.533 0.369 0.438
0.271 0.317 0.415 0.968 0.965 0.589 0.324 0.419
0.271 0.317 0.419 0.968 0.965 0.586 0.324 0.420
token_length_mean
token_length_median
token_length_std
n_tokens
n_unique_tokens
proportion_unique_tokens
mean_word_length
alpha_ratio
1.0
0.8
0.6
0.4
0.2
0.0

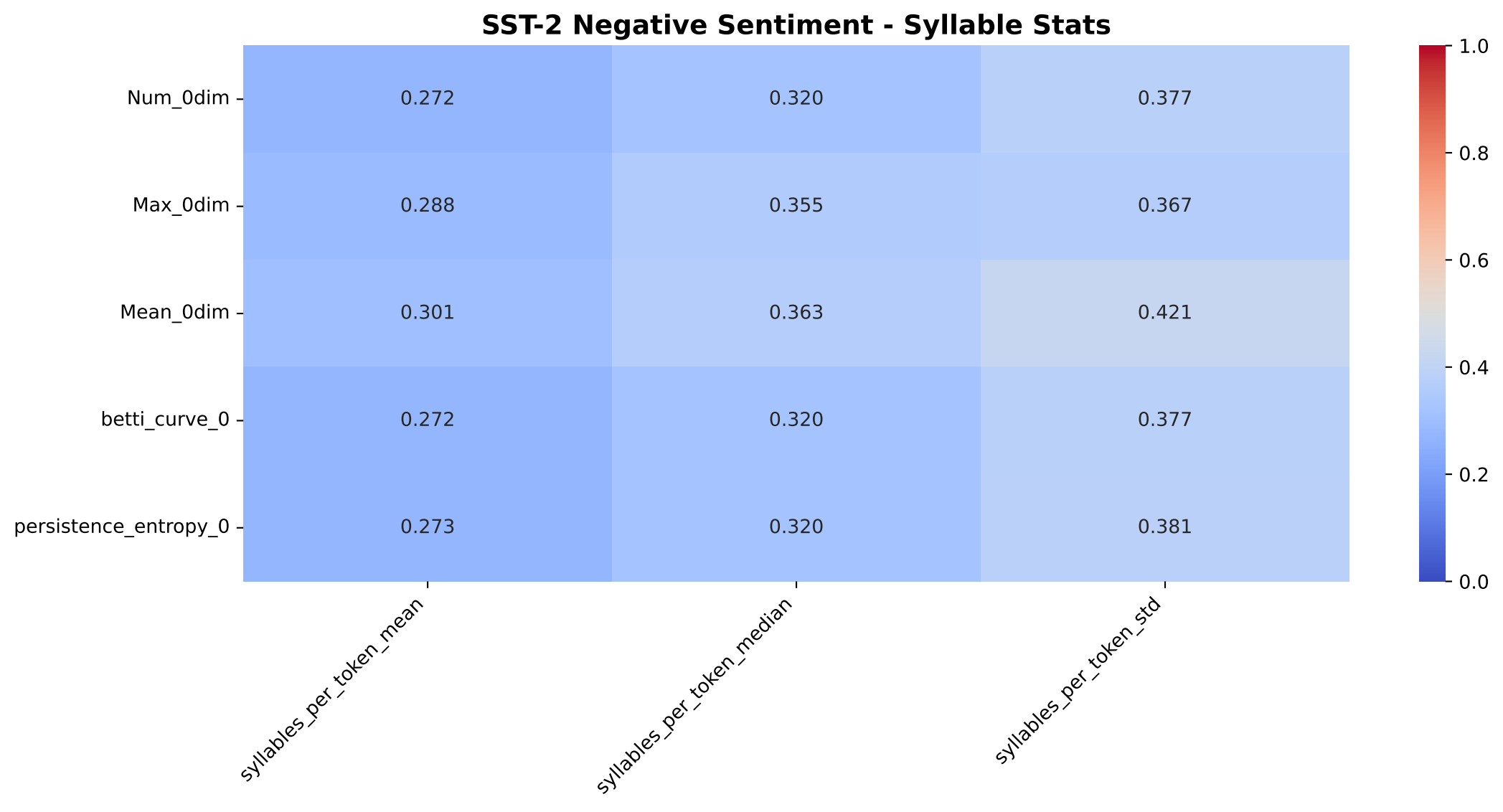

SST-2 Difference: Positive - Negative - Token Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.021 0.016 -0.048 0.004 0.003 0.014 0.001 0.009
0.008 -0.005 -0.064 -0.047 -0.042 -0.064 -0.007 0.014
0.026 0.022 -0.070 -0.010 -0.010 0.004 0.011 0.018
0.021 0.016 -0.048 0.004 0.003 0.014 0.001 0.009
0.022 0.017 -0.049 0.004 0.002 0.015 0.002 0.009
token_length_mean
token_length_median
token_length_std
n_tokens
n_unique_tokens
proportion_unique_tokens
mean_word_length
alpha_ratio
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

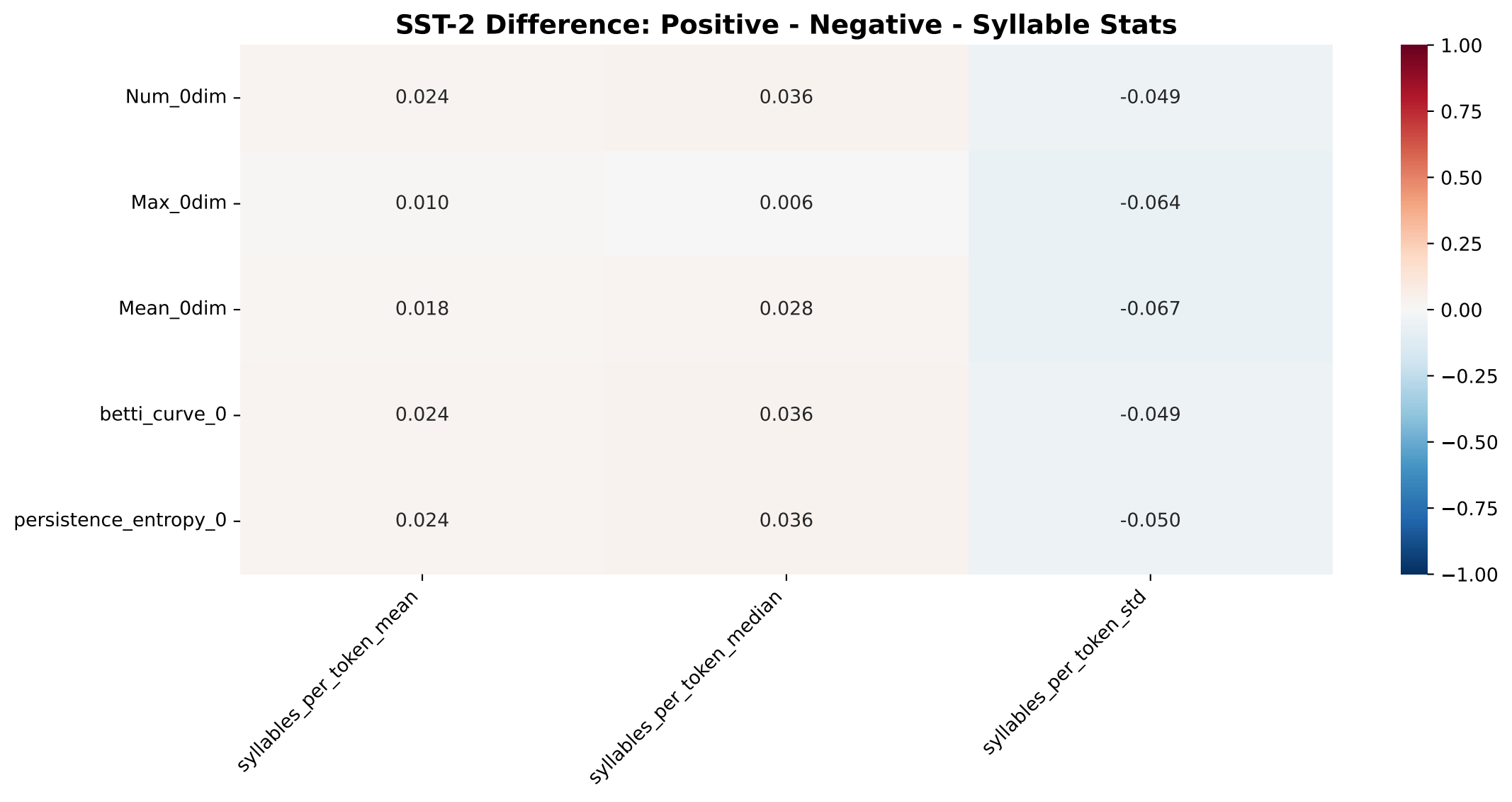

SST-2 Positive Sentiment - Sentence Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.959
0.956
0.171
0.182
0.983
0.836
0.833
0.141
0.150
0.859
0.898
0.895
0.140
0.145
0.915
0.959
0.956
0.171
0.182
0.983
0.959
0.955
0.167
0.178
0.982
sentence_length_mean
sentence_length_median
sentence_length_std
n_sentences
doc_length
1.0
0.8
0.6
0.4
0.2
0.0

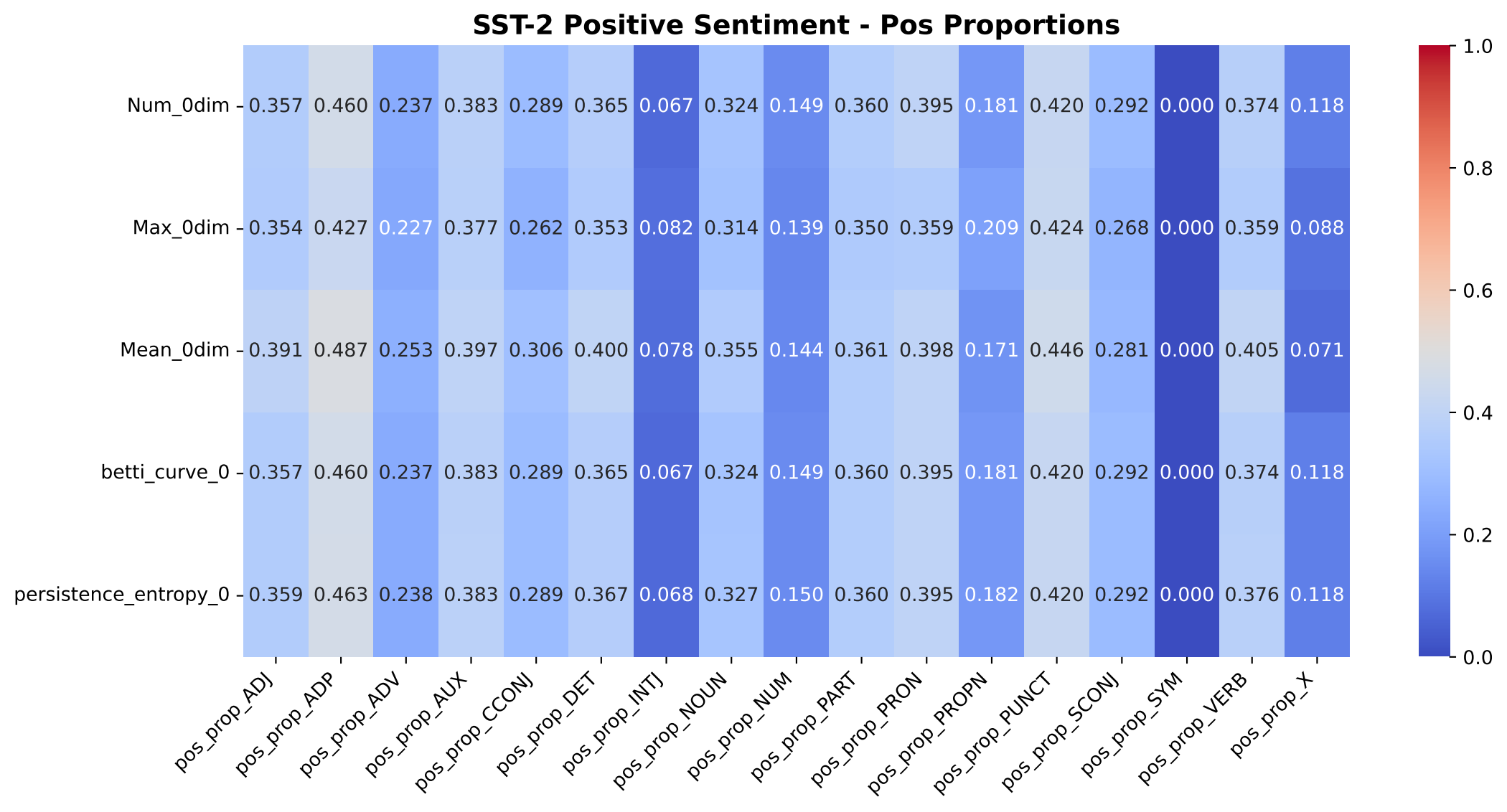

SST-2 Negative Sentiment - Sentence Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.964
0.964
0.100
0.086
0.982
0.890
0.890
0.079
0.068
0.904
0.912
0.912
0.072
0.064
0.926
0.964
0.964
0.100
0.086
0.982
0.964
0.964
0.099
0.086
0.981
sentence_length_mean
sentence_length_median
sentence_length_std
n_sentences
doc_length
1.0
0.8
0.6
0.4
0.2
0.0

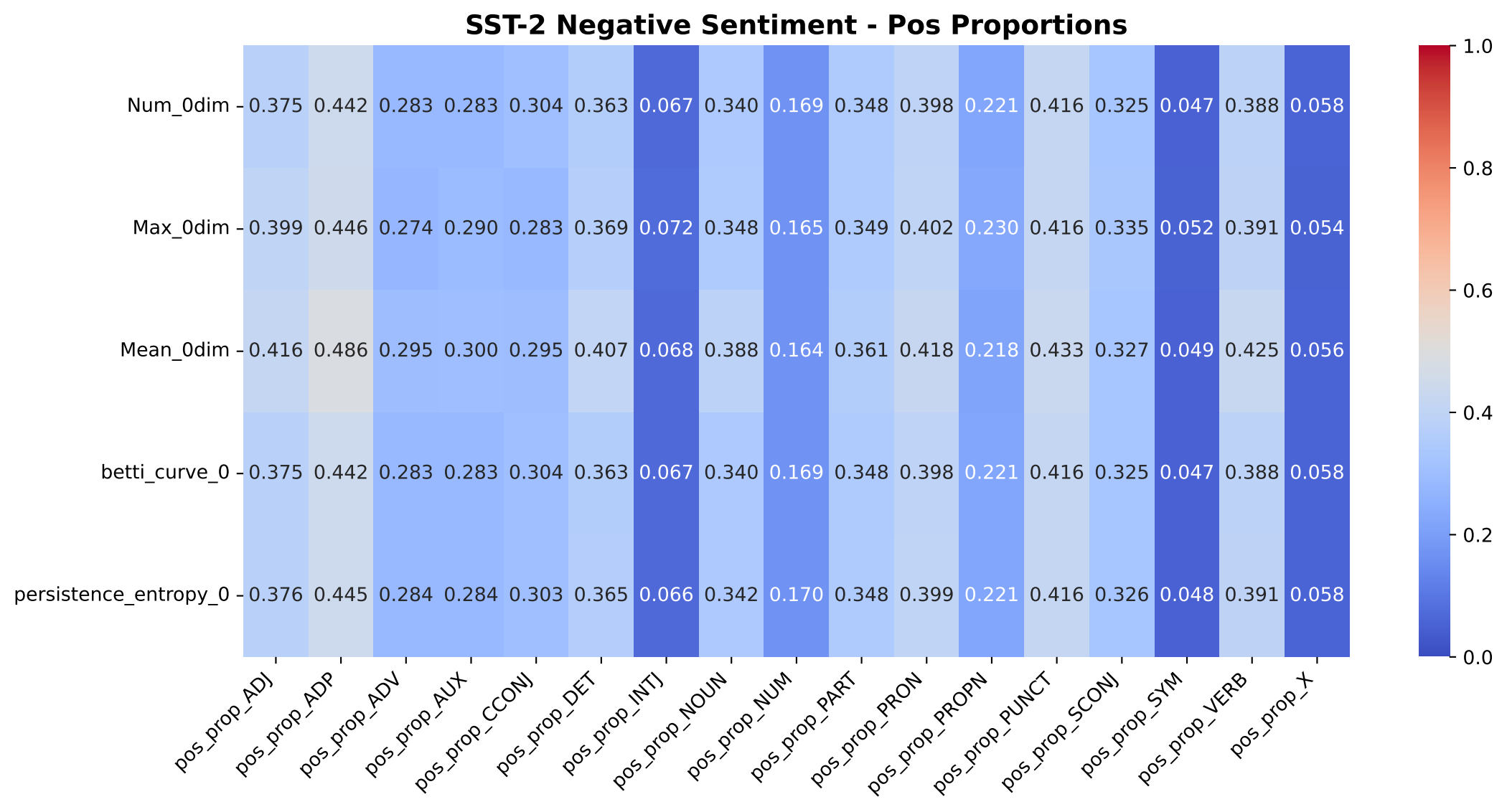

SST-2 Difference: Positive - Negative - Sentence Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
-0.005
-0.008
0.071
0.096
0.001
-0.054
-0.057
0.062
0.083
-0.046
-0.014
-0.017
0.068
0.081
-0.011
-0.005
-0.008
0.071
0.096
0.001
-0.005
-0.008
0.068
0.092
0.000
sentence_length_mean
sentence_length_median
sentence_length_std
n_sentences
doc_length
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

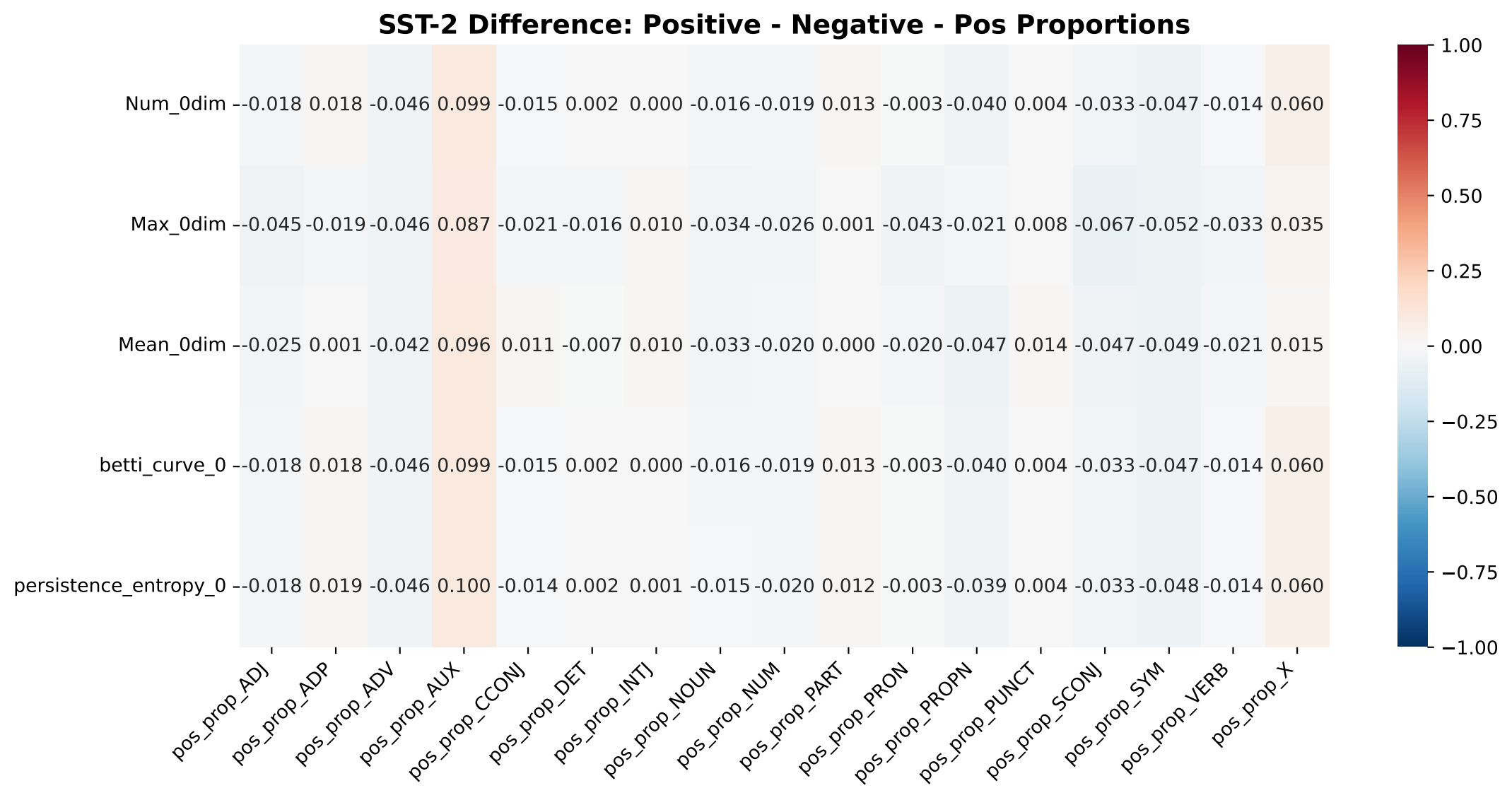

SST-2 Positive Sentiment - Syllable Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.297
0.356
0.328
0.298
0.361
0.303
0.319
0.392
0.354
0.297
0.356
0.328
0.298
0.356
0.331
syllables_per_token_mean
syllables_per_token_median
syllables_per_token_std
1.0
0.8
0.6
0.4
0.2
0.0

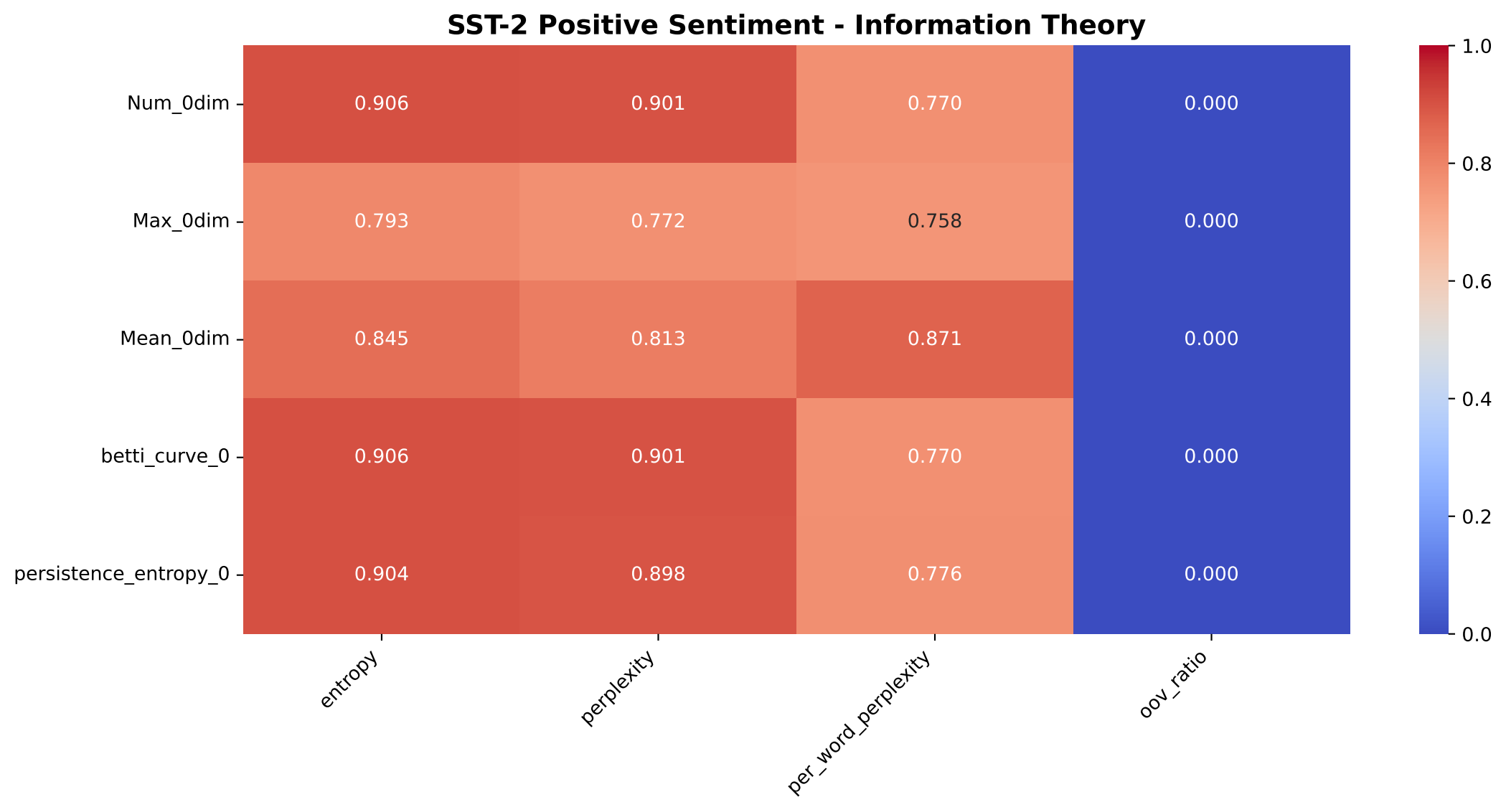

SST-2 Negative Sentiment - Syllable Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.272
0.320
0.377
0.288
0.355
0.367
0.301
0.363
0.421
0.272
0.320
0.377
0.273
0.320
0.381
syllables_per_token_mean
syllables_per_token_median
syllables_per_token_std
1.0
0.8
0.6
0.4
0.2
0.0

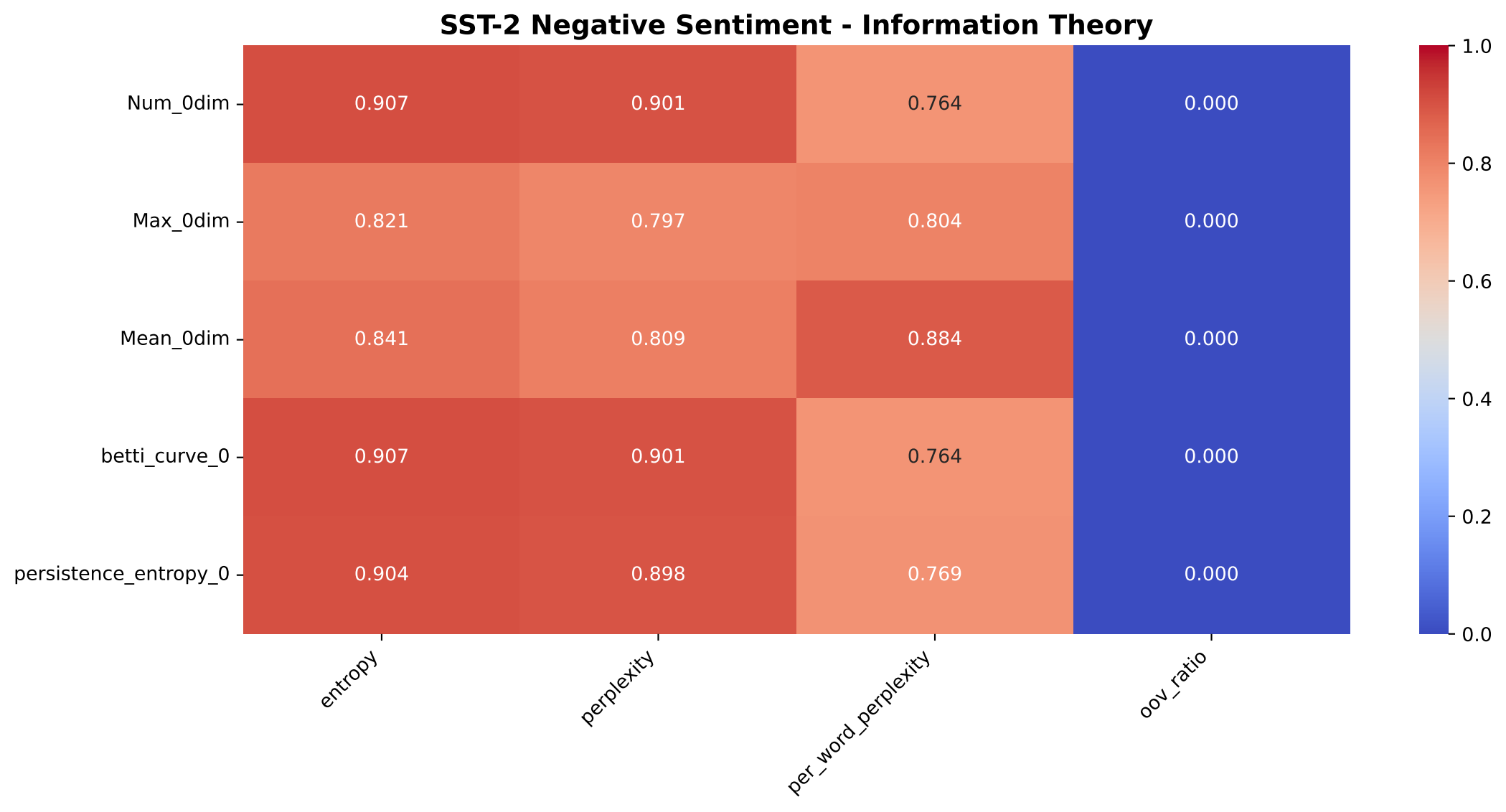

SST-2 Difference: Positive - Negative - Syllable Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.024
0.036
-0.049
0.010
0.006
-0.064
0.018
0.028
-0.067
0.024
0.036
-0.049
0.024
0.036
-0.050
syllables_per_token_mean
syllables_per_token_median
syllables_per_token_std
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

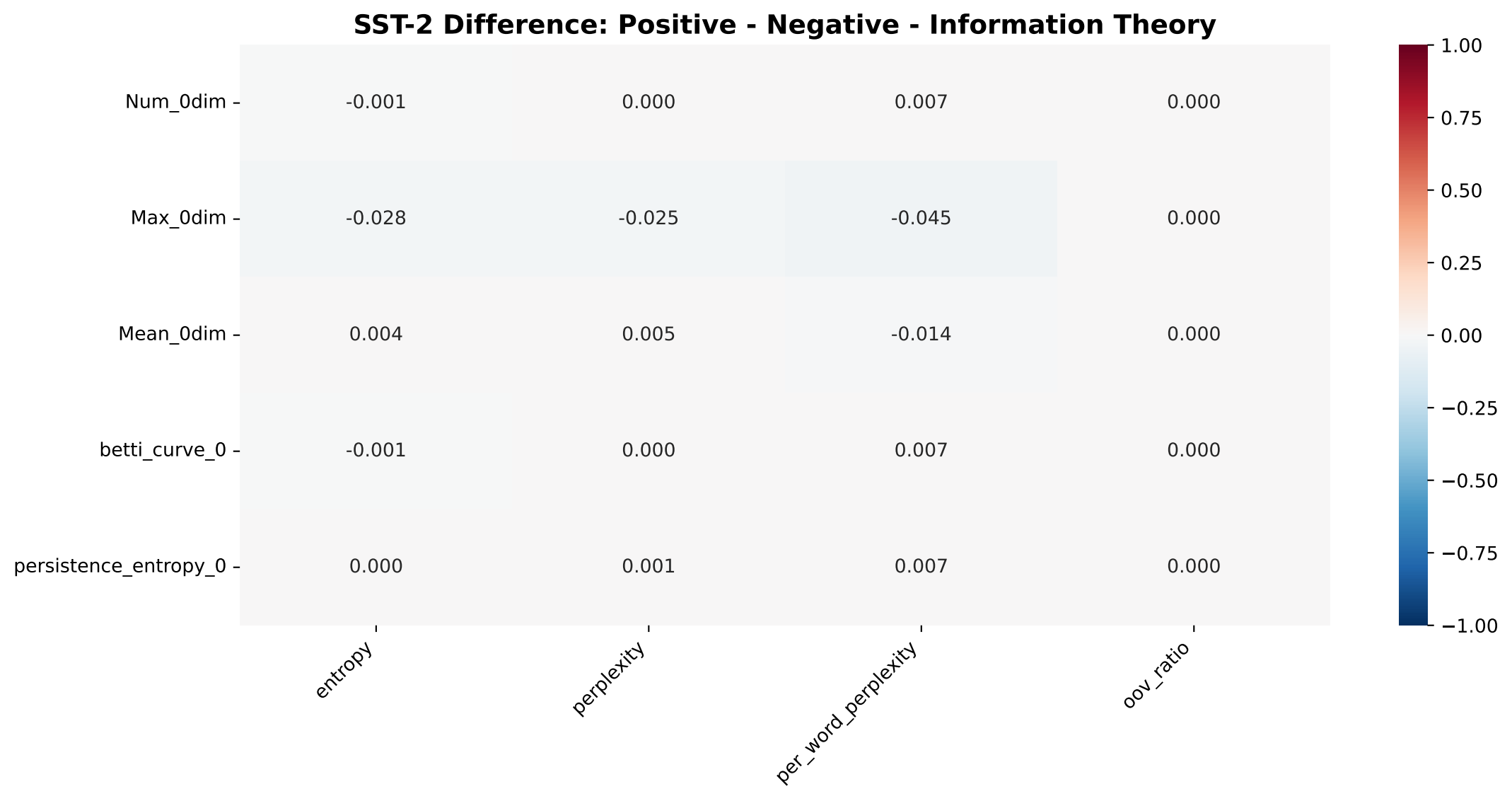

SST-2 Positive Sentiment - Pos Proportions
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.357 0.460 0.237 0.383 0.289 0.365 0.067 0.324 0.149 0.360 0.395 0.181 0.420 0.292 0.000 0.374 0.118
0.354 0.427 0.227 0.377 0.262 0.353 0.082 0.314 0.139 0.350 0.359 0.209 0.424 0.268 0.000 0.359 0.088
0.391 0.487 0.253 0.397 0.306 0.400 0.078 0.355 0.144 0.361 0.398 0.171 0.446 0.281 0.000 0.405 0.071
0.357 0.460 0.237 0.383 0.289 0.365 0.067 0.324 0.149 0.360 0.395 0.181 0.420 0.292 0.000 0.374 0.118
0.359 0.463 0.238 0.383 0.289 0.367 0.068 0.327 0.150 0.360 0.395 0.182 0.420 0.292 0.000 0.376 0.118
pos_prop_ADJ
pos_prop_ADP
pos_prop_ADV
pos_prop_AUX
pos_prop_CCONJ
pos_prop_DET
pos_prop_INTJ
pos_prop_NOUN
pos_prop_NUM
pos_prop_PART
pos_prop_PRON
pos_prop_PROPN
pos_prop_PUNCT
pos_prop_SCONJ
pos_prop_SYM
pos_prop_VERB
pos_prop_X
1.0
0.8
0.6
0.4
0.2
0.0

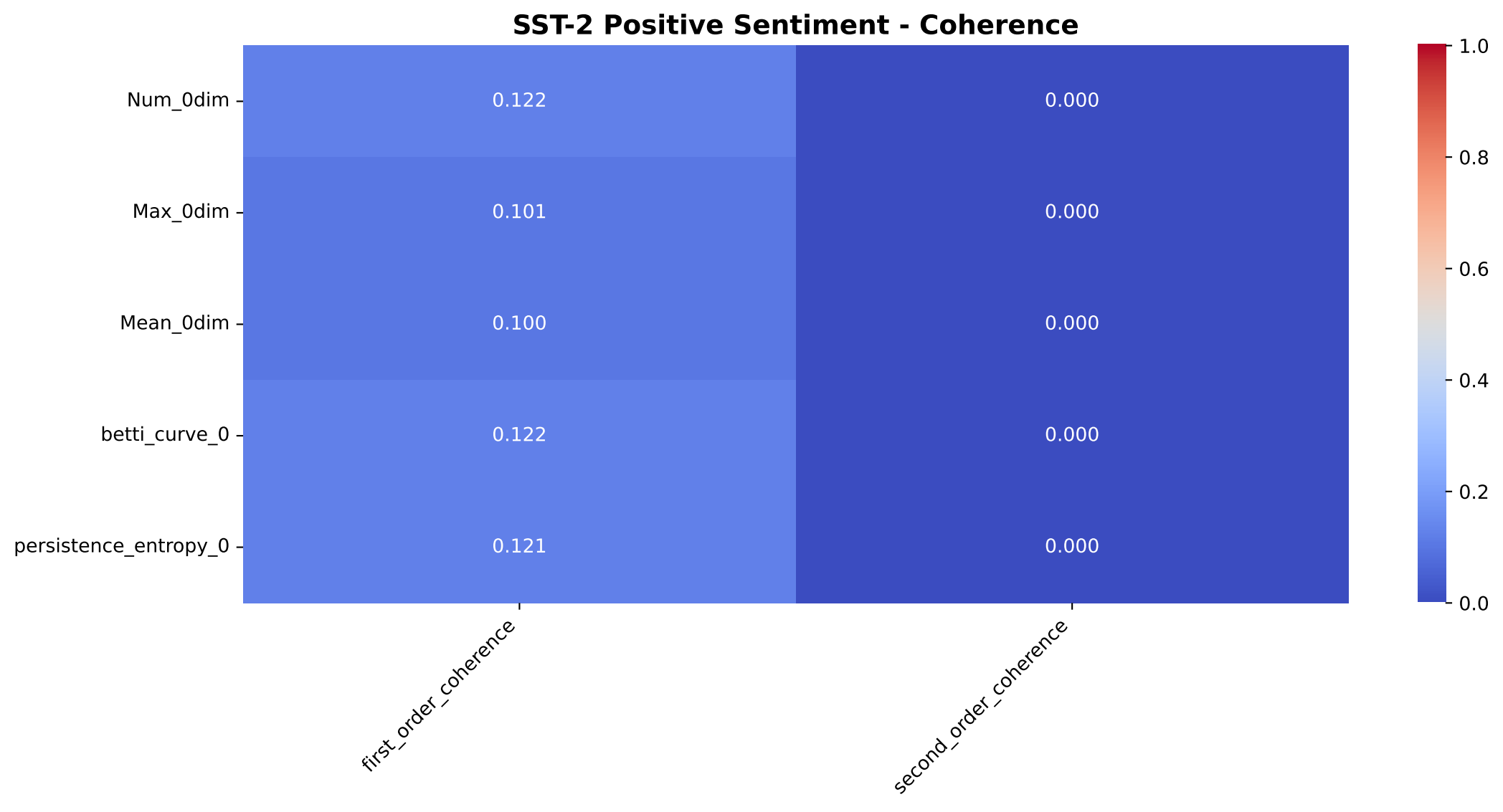

SST-2 Negative Sentiment - Pos Proportions
Num_0dim 0.375 0.442 0.283 0.283 0.304 0.363 0.067 0.340 0.169 0.348 0.398 0.221 0.416 0.325 0.047 0.388 0.058
Max_0dim 0.399 0.446 0.274 0.290 0.283 0.369 0.072 0.348 0.165 0.349 0.402 0.230 0.416 0.335 0.052 0.391 0.054
Mean_0dim 0.416 0.486 0.295 0.300 0.295 0.407 0.068 0.388 0.164 0.361 0.418 0.218 0.433 0.327 0.049 0.425 0.056
betti_curve_0 0.375 0.442 0.283 0.283 0.304 0.363 0.067 0.340 0.169 0.348 0.398 0.221 0.416 0.325 0.047 0.388 0.058
persistence_entropy_0 0.376 0.445 0.284 0.284 0.303 0.365 0.066 0.342 0.170 0.348 0.399 0.221 0.416 0.326 0.048 0.391 0.058
pos_prop_ADJ
pos_prop_ADP
pos_prop_ADV
pos_prop_AUX
pos_prop_CCONJ
pos_prop_DET
pos_prop_INTJ
pos_prop_NOUN
pos_prop_NUM
pos_prop_PART
pos_prop_PRON
pos_prop_PROPN
pos_prop_PUNCT
pos_prop_SCONJ
pos_prop_SYM
pos_prop_VERB
pos_prop_X
1.0
0.8
0.6
0.4
0.2
0.0

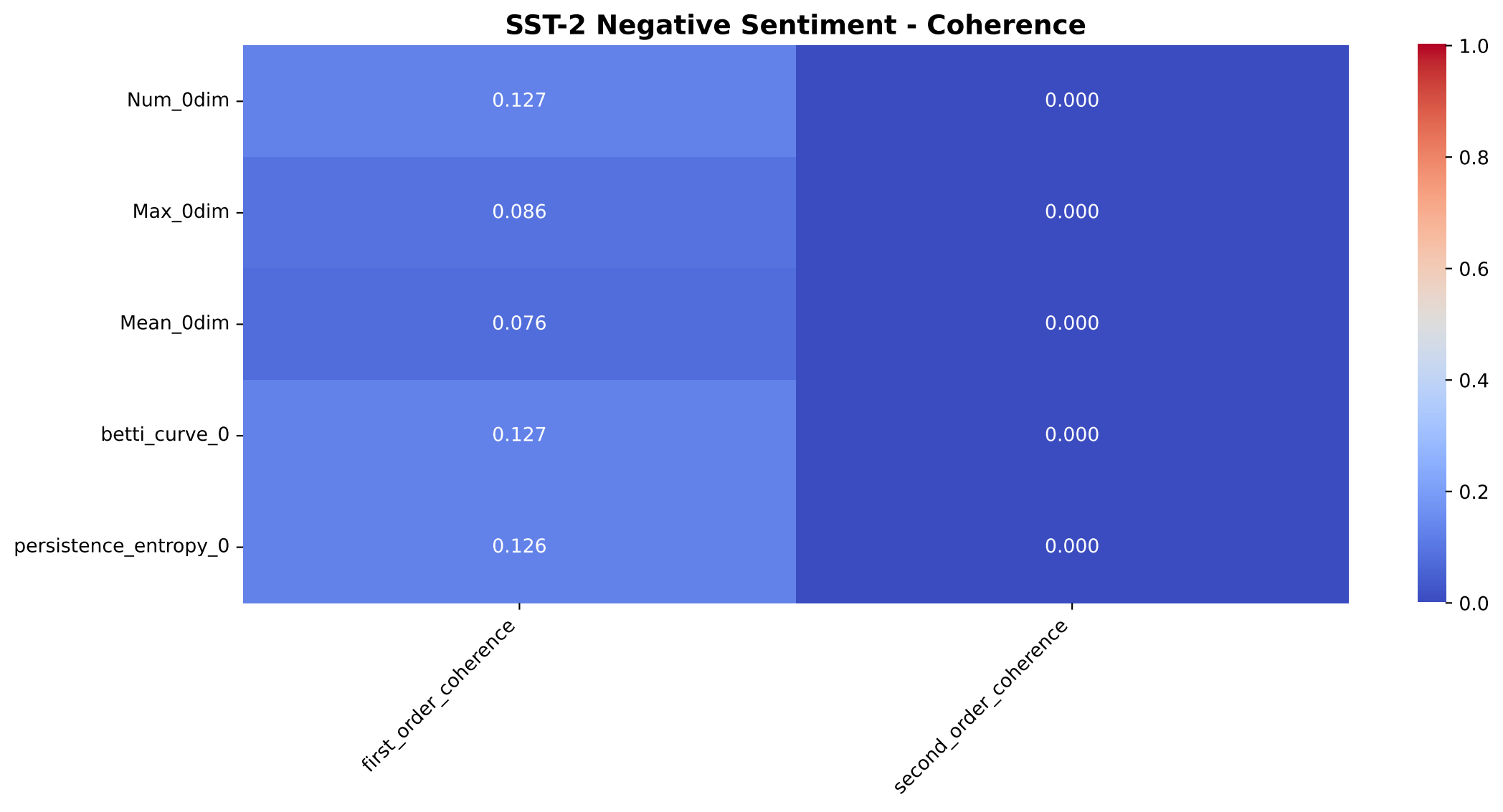

SST-2 Difference: Positive - Negative - Pos Proportions
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
-0.018 0.018 -0.046 0.099 -0.015 0.002 0.000 -0.016 -0.019 0.013 -0.003 -0.040 0.004 -0.033 -0.047 -0.014 0.060
-0.045 -0.019 -0.046 0.087 -0.021 -0.016 0.010 -0.034 -0.026 0.001 -0.043 -0.021 0.008 -0.067 -0.052 -0.033 0.035
-0.025 0.001 -0.042 0.096 0.011 -0.007 0.010 -0.033 -0.020 0.000 -0.020 -0.047 0.014 -0.047 -0.049 -0.021 0.015
-0.018 0.018 -0.046 0.099 -0.015 0.002 0.000 -0.016 -0.019 0.013 -0.003 -0.040 0.004 -0.033 -0.047 -0.014 0.060
-0.018 0.019 -0.046 0.100 -0.014 0.002 0.001 -0.015 -0.020 0.012 -0.003 -0.039 0.004 -0.033 -0.048 -0.014 0.060
pos_prop_ADJ
pos_prop_ADP
pos_prop_ADV
pos_prop_AUX
pos_prop_CCONJ
pos_prop_DET
pos_prop_INTJ
pos_prop_NOUN
pos_prop_NUM
pos_prop_PART
pos_prop_PRON
pos_prop_PROPN
pos_prop_PUNCT
pos_prop_SCONJ
pos_prop_SYM
pos_prop_VERB
pos_prop_X
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

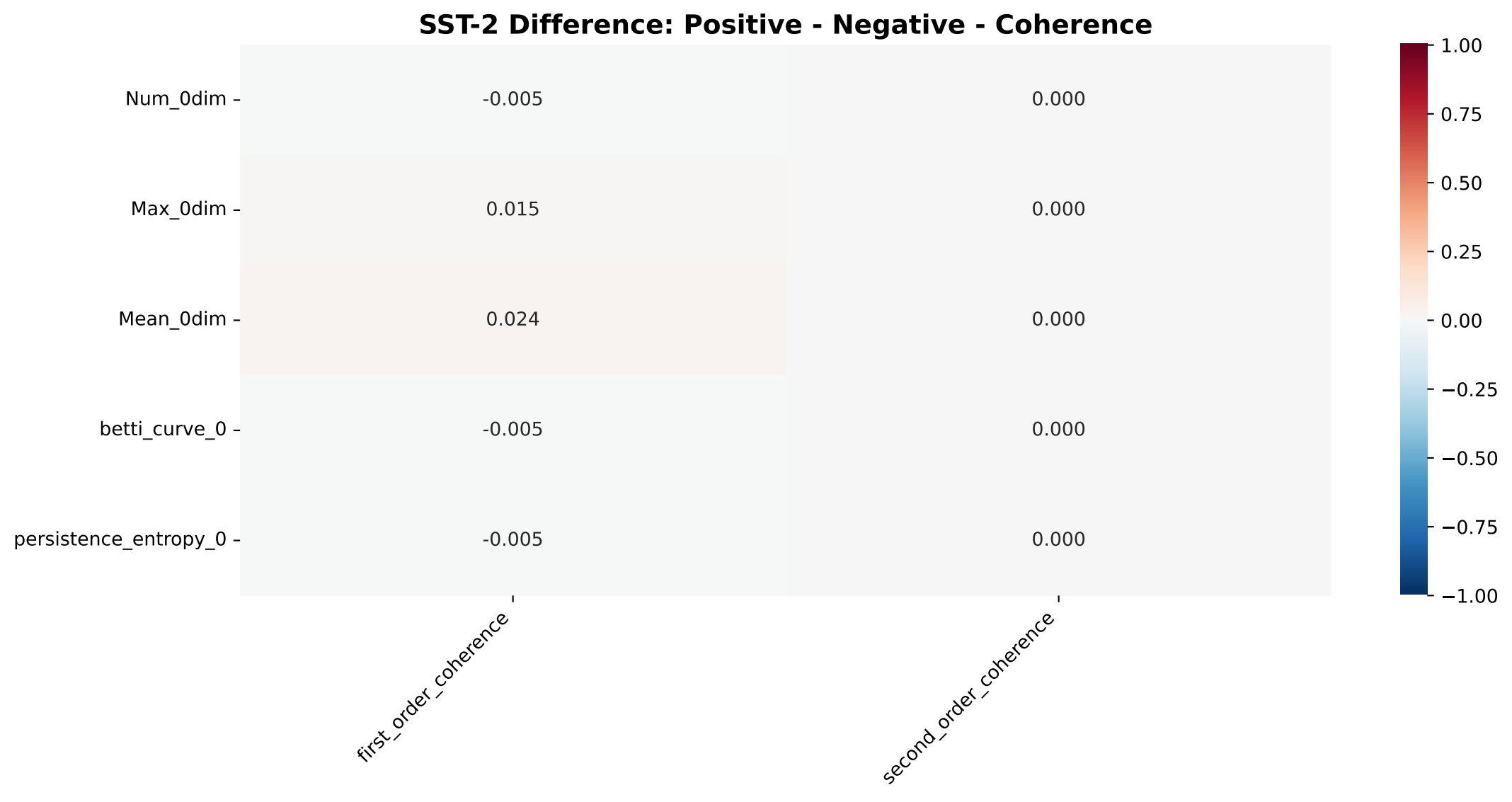

SST-2 Positive Sentiment - Information Theory
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.906
0.901
0.770
0.000
0.793
0.772
0.758
0.000
0.845
0.813
0.871
0.000
0.906
0.901
0.770
0.000
0.904
0.898
0.776
0.000
entropy
perplexity
per_word_perplexity
oov_ratio
1.0
0.8
0.6
0.4
0.2
0.0

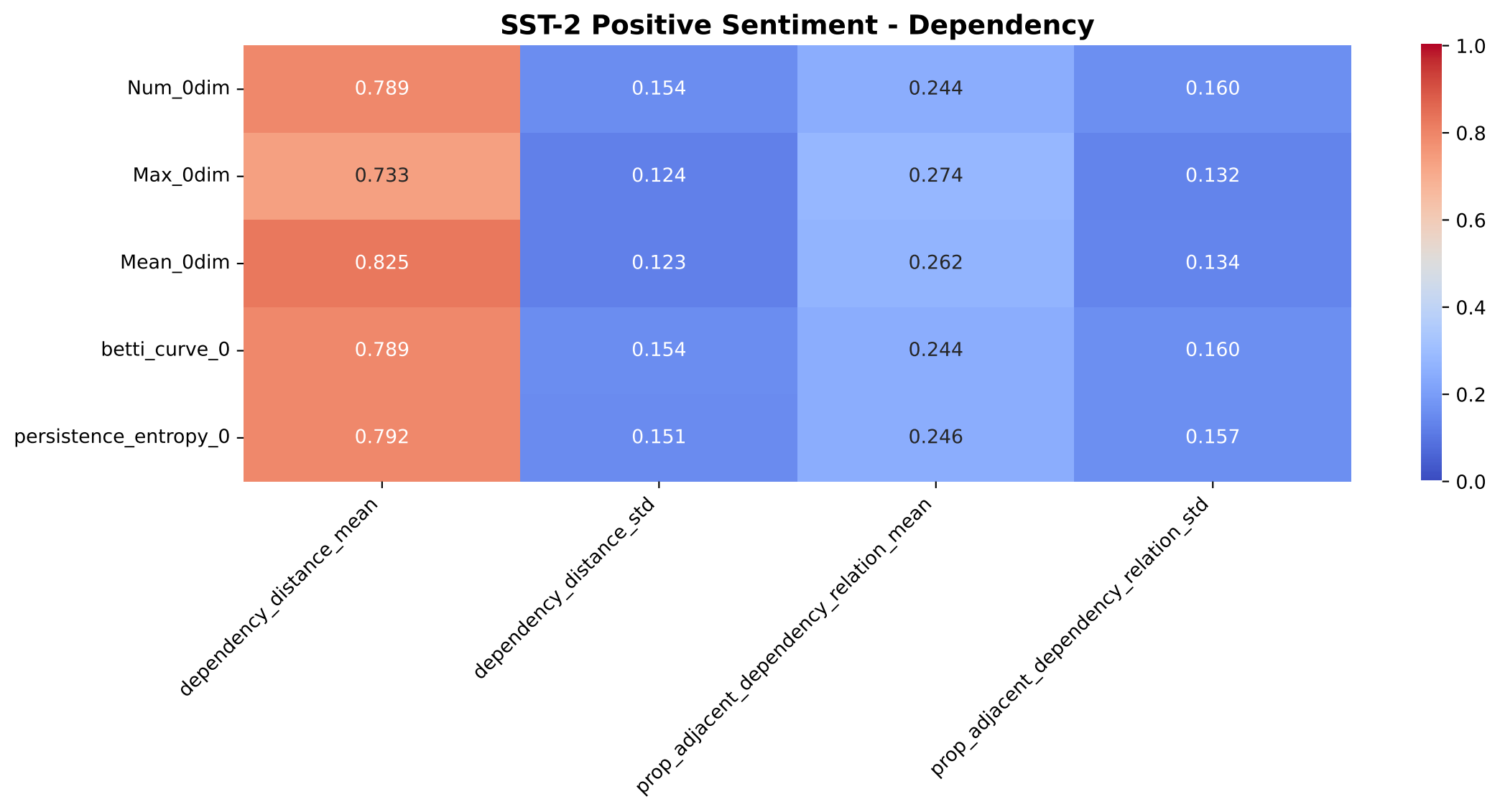

SST-2 Negative Sentiment - Information Theory
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.907
0.901
0.764
0.000
0.821
0.797
0.804
0.000
0.841
0.809
0.884
0.000
0.907
0.901
0.764
0.000
0.904
0.898
0.769
0.000
entropy
perplexity
per_word_perplexity
oov_ratio
1.0
0.8
0.6
0.4
0.2
0.0

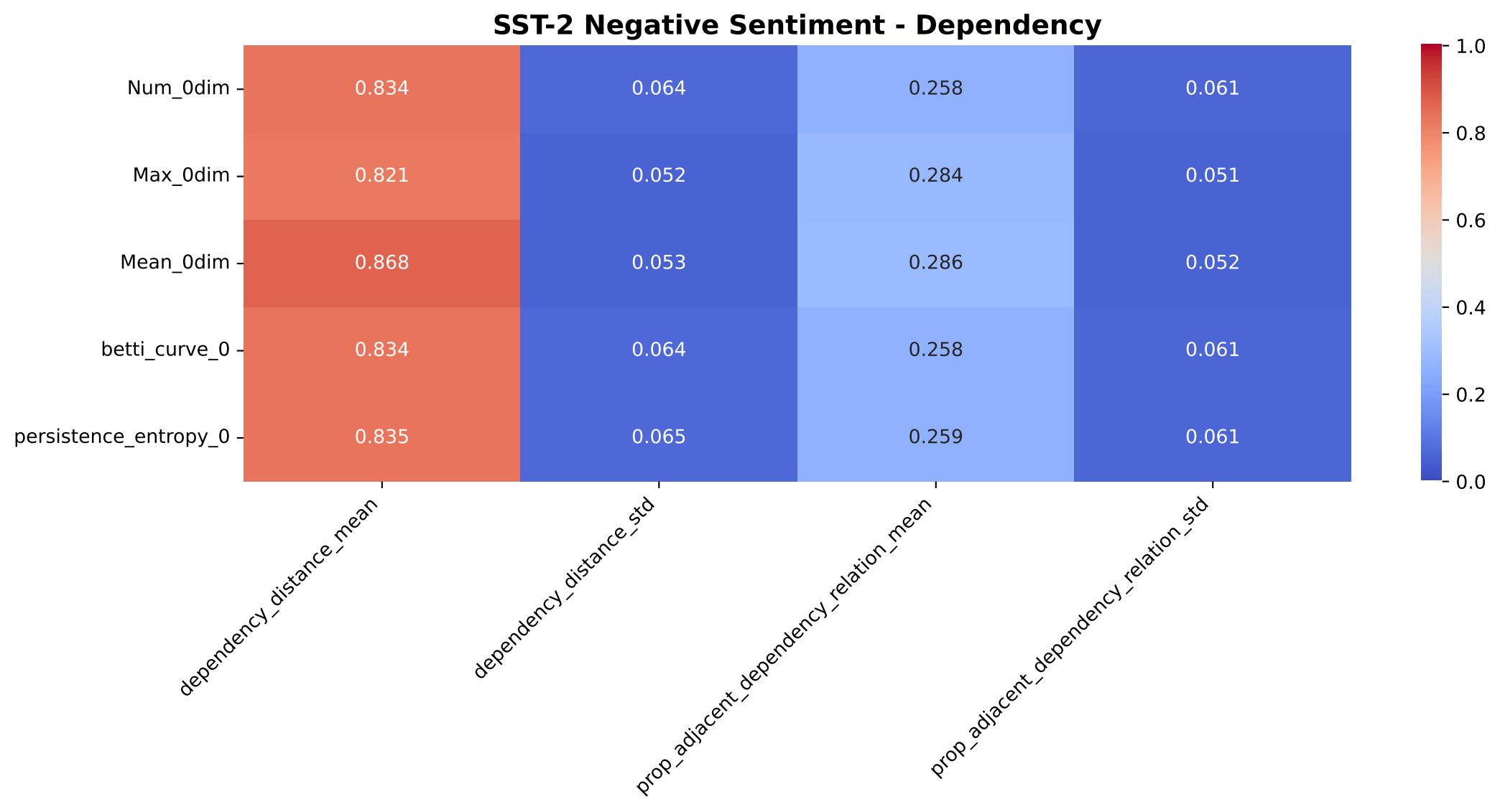

SST-2 Difference: Positive - Negative - Information Theory
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
-0.001
0.000
0.007
0.000
-0.028
-0.025
-0.045
0.000
0.004
0.005
-0.014
0.000
-0.001
0.000
0.007
0.000
0.000
0.001
0.007
0.000
entropy
perplexity
per_word_perplexity
oov_ratio
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

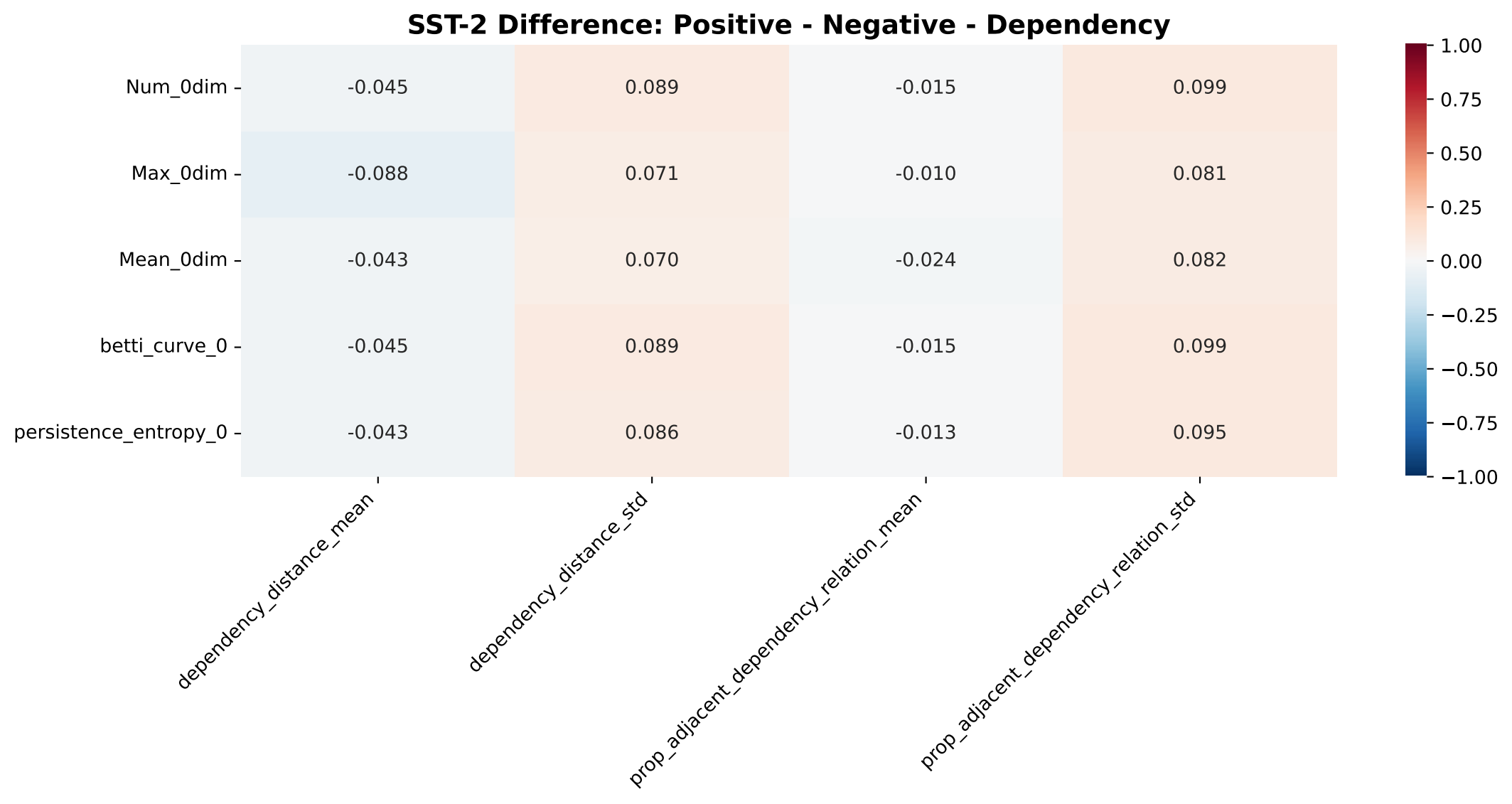

SST-2 Positive Sentiment - Coherence
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.122
0.000
0.101
0.000
0.100
0.000
0.122
0.000
0.121
0.000
first_order_coherence
second_order_coherence
1.0
0.8
0.6
0.4
0.2
0.0

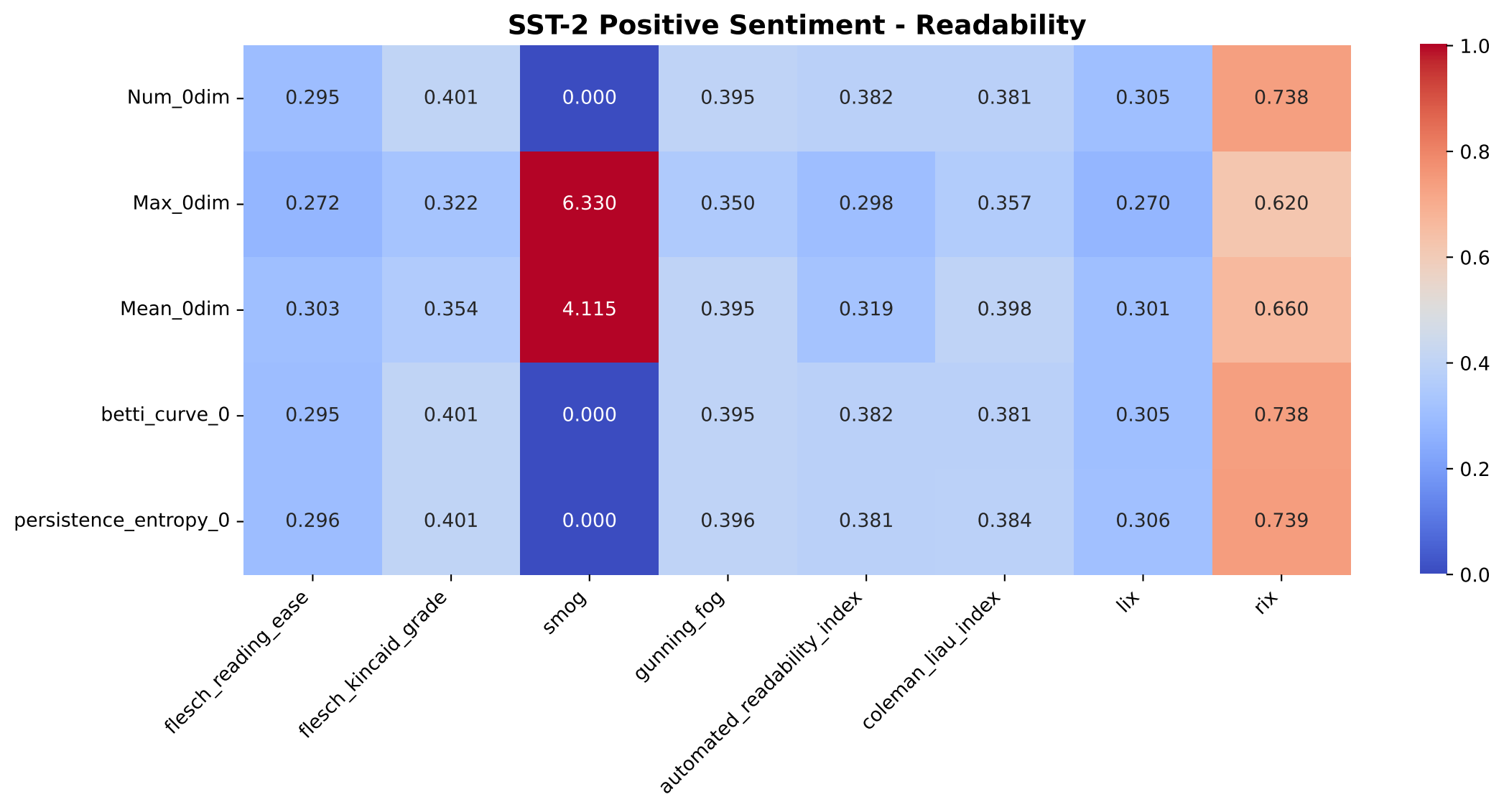

SST-2 Negative Sentiment - Coherence
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.127
0.000
0.086
0.000
0.076
0.000
0.127
0.000
0.126
0.000
first_order_coherence
second_order_coherence
1.0
0.8
0.6
0.4
0.2
0.0

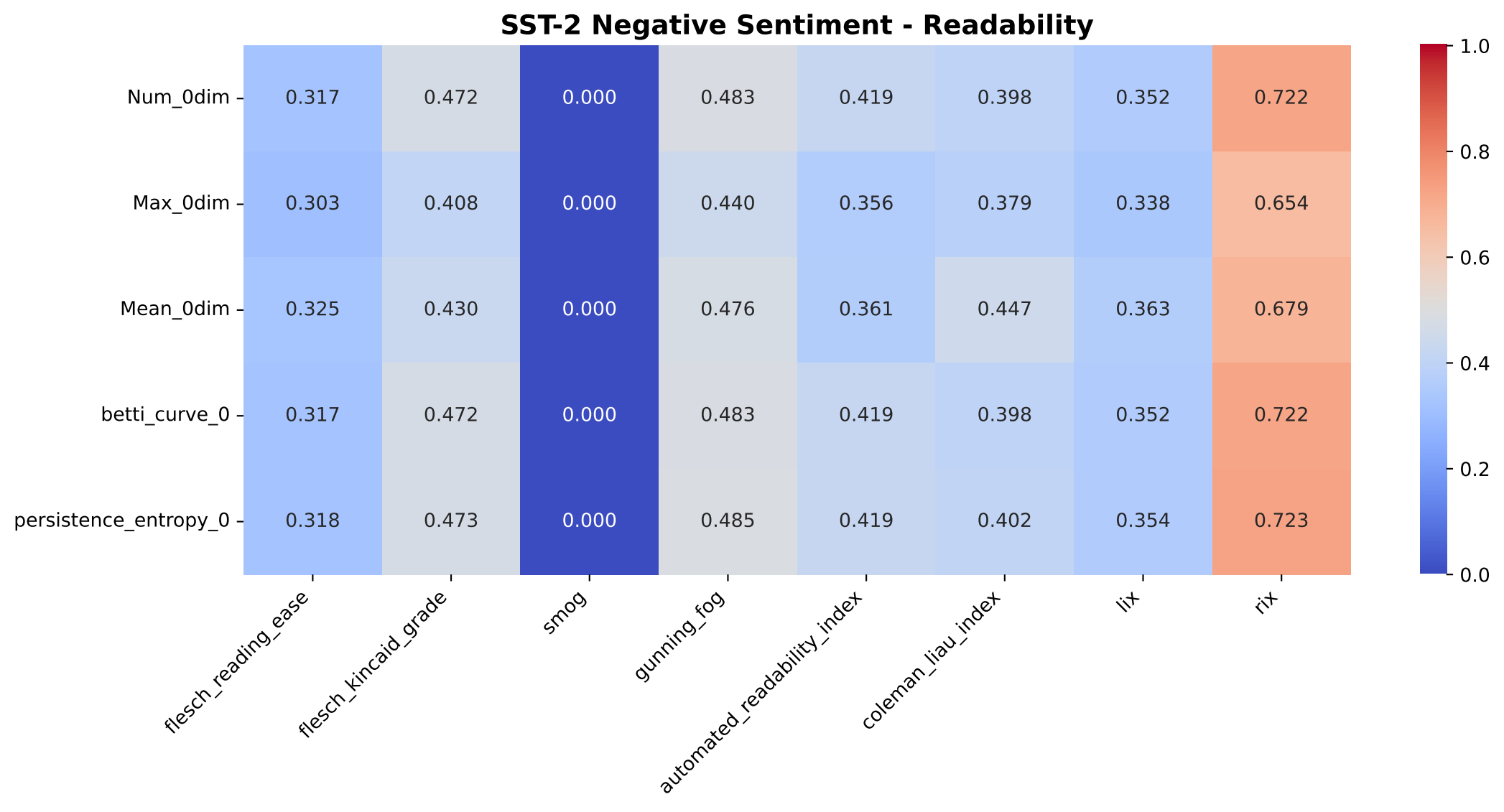

SST-2 Difference: Positive - Negative - Coherence
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
-0.005
0.000
0.015
0.000
0.024
0.000
-0.005
0.000
-0.005
0.000
first_order_coherence
second_order_coherence
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

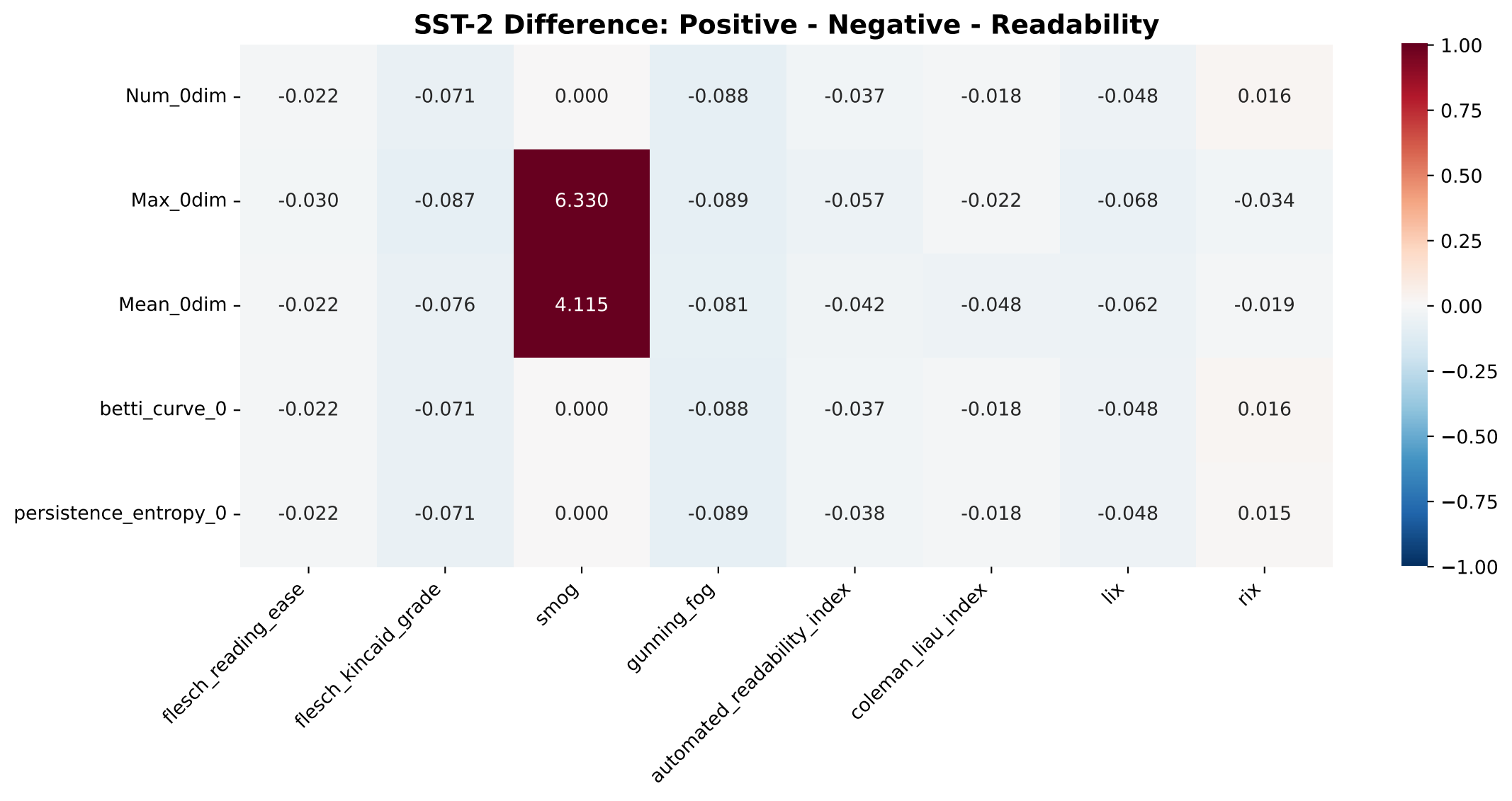

SST-2 Positive Sentiment - Dependency
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.789
0.154
0.244
0.160
0.733
0.124
0.274
0.132
0.825
0.123
0.262
0.134
0.789
0.154
0.244
0.160
0.792
0.151
0.246
0.157
dependency_distance_mean
dependency_distance_std
prop_adjacent_dependency_relation_mean
prop_adjacent_dependency_relation_std
1.0
0.8
0.6
0.4
0.2
0.0

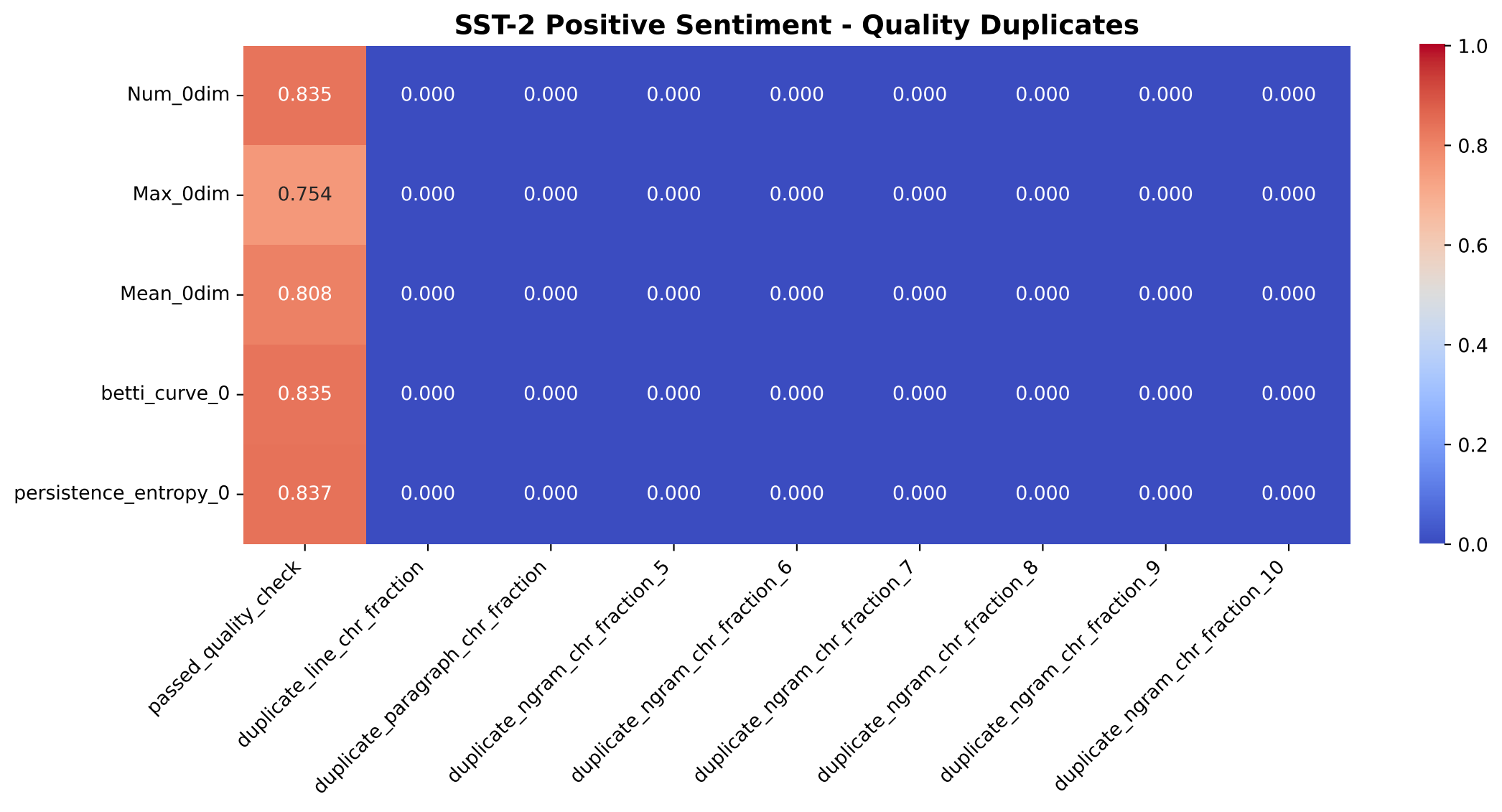

SST-2 Negative Sentiment - Dependency
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.834
0.064
0.258
0.061
0.821
0.052
0.284
0.051
0.868
0.053
0.286
0.052
0.834
0.064
0.258
0.061
0.835
0.065
0.259
0.061
dependency_distance_mean
dependency_distance_std
prop_adjacent_dependency_relation_mean
prop_adjacent_dependency_relation_std
1.0
0.8
0.6
0.4
0.2
0.0

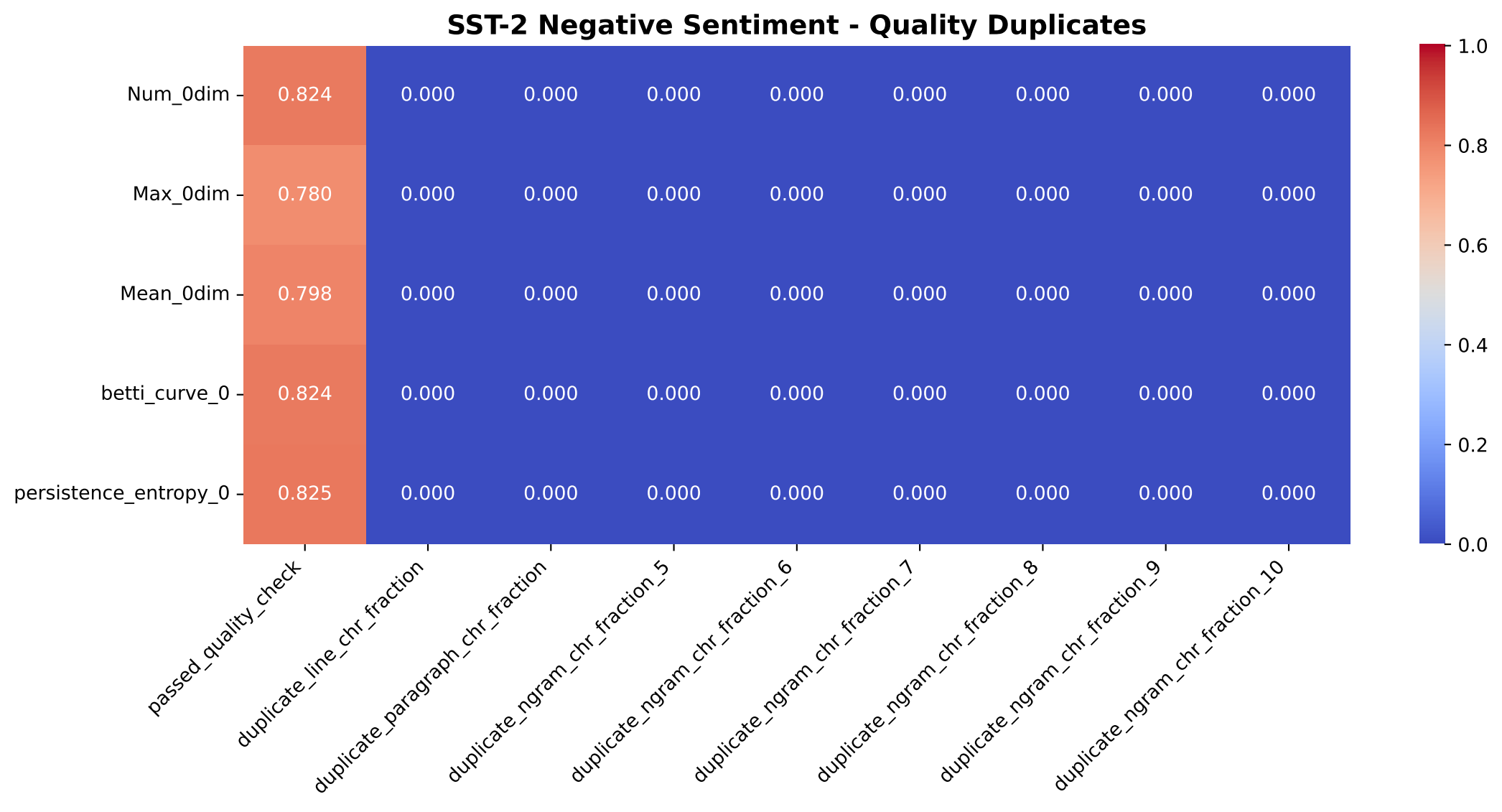

SST-2 Difference: Positive - Negative - Dependency
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
-0.045
0.089
-0.015
0.099
-0.088
0.071
-0.010
0.081
-0.043
0.070
-0.024
0.082
-0.045
0.089
-0.015
0.099
-0.043
0.086
-0.013
0.095
dependency_distance_mean
dependency_distance_std
prop_adjacent_dependency_relation_mean
prop_adjacent_dependency_relation_std
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

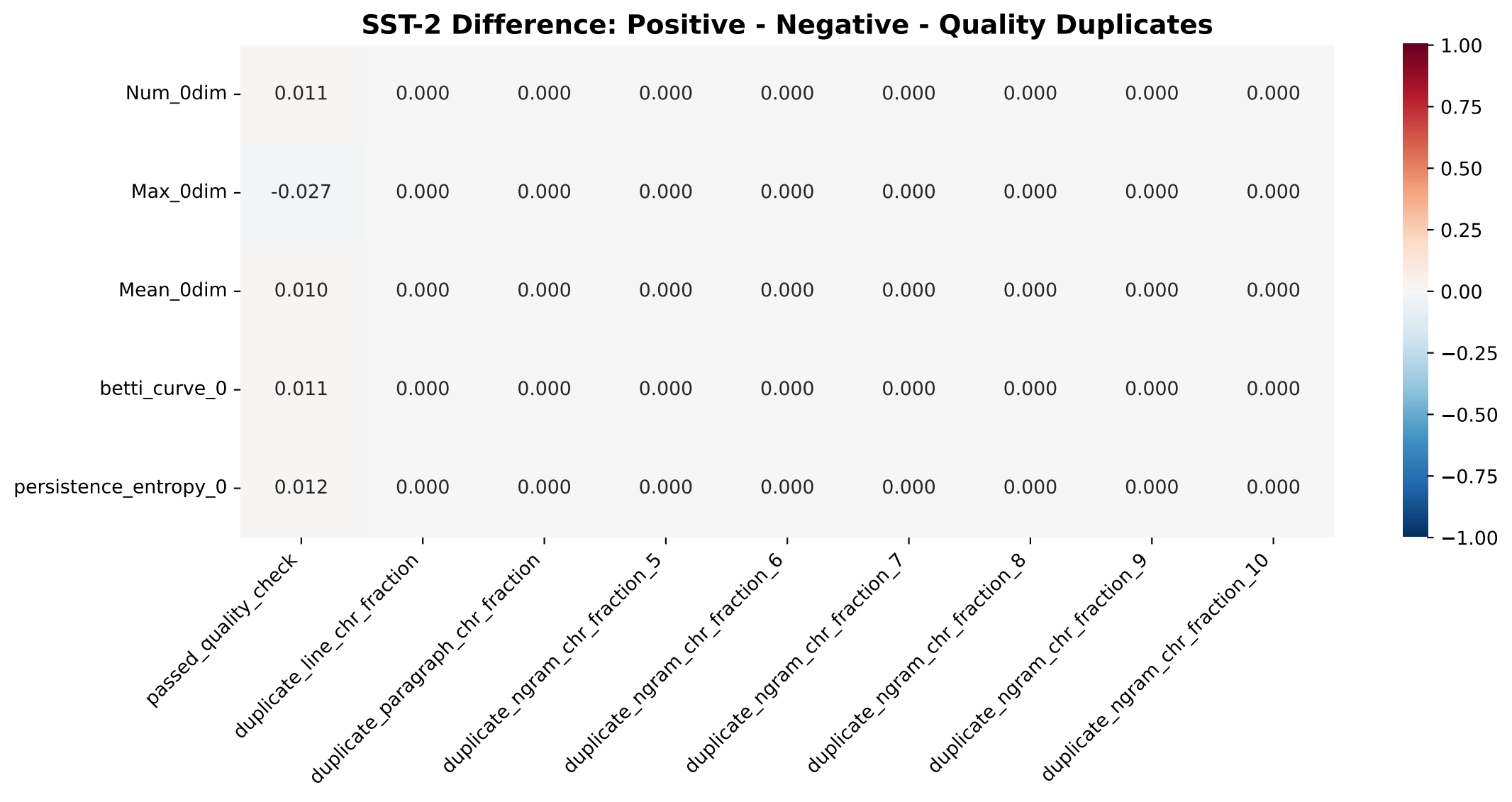

SST-2 Positive Sentiment - Readability
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.295 0.401 0.000 0.395 0.382 0.381 0.305 0.738
0.272 0.322 6.330 0.350 0.298 0.357 0.270 0.620
0.303 0.354 4.115 0.395 0.319 0.398 0.301 0.660
0.295 0.401 0.000 0.395 0.382 0.381 0.305 0.738
0.296 0.401 0.000 0.396 0.381 0.384 0.306 0.739
flesch_reading_ease
flesch_kincaid_grade
smog
gunning_fog
automated_readability_index
coleman_liau_index
lix
rix
1.0
0.8
0.6
0.4
0.2
0.0

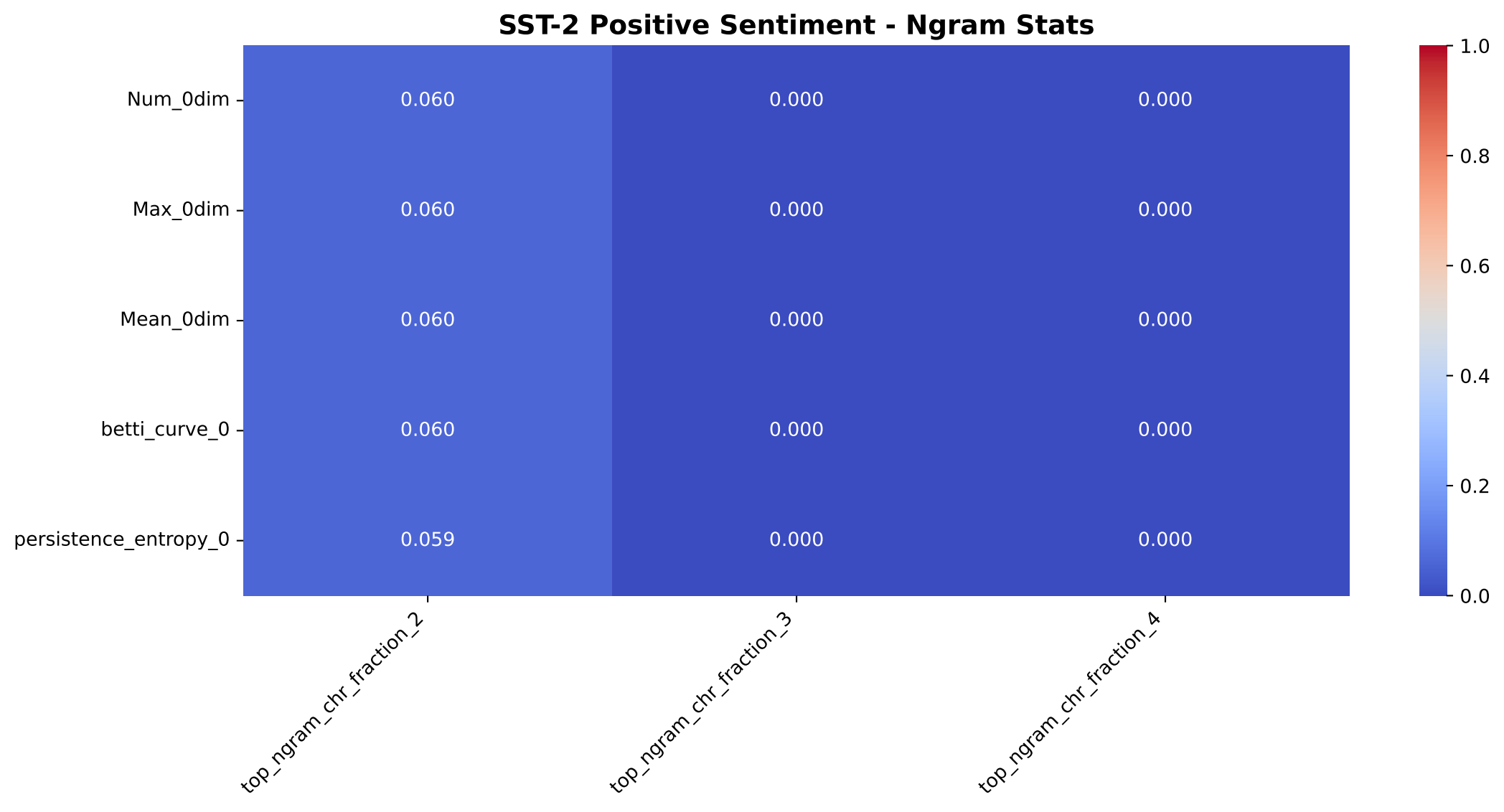

SST-2 Negative Sentiment - Readability
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.317 0.472 0.000 0.483 0.419 0.398 0.352 0.722
0.303 0.408 0.000 0.440 0.356 0.379 0.338 0.654
0.325 0.430 0.000 0.476 0.361 0.447 0.363 0.679
0.317 0.472 0.000 0.483 0.419 0.398 0.352 0.722
0.318 0.473 0.000 0.485 0.419 0.402 0.354 0.723
flesch_reading_ease
flesch_kincaid_grade
smog
gunning_fog
automated_readability_index
coleman_liau_index
lix
rix
1.0
0.8
0.6
0.4
0.2
0.0

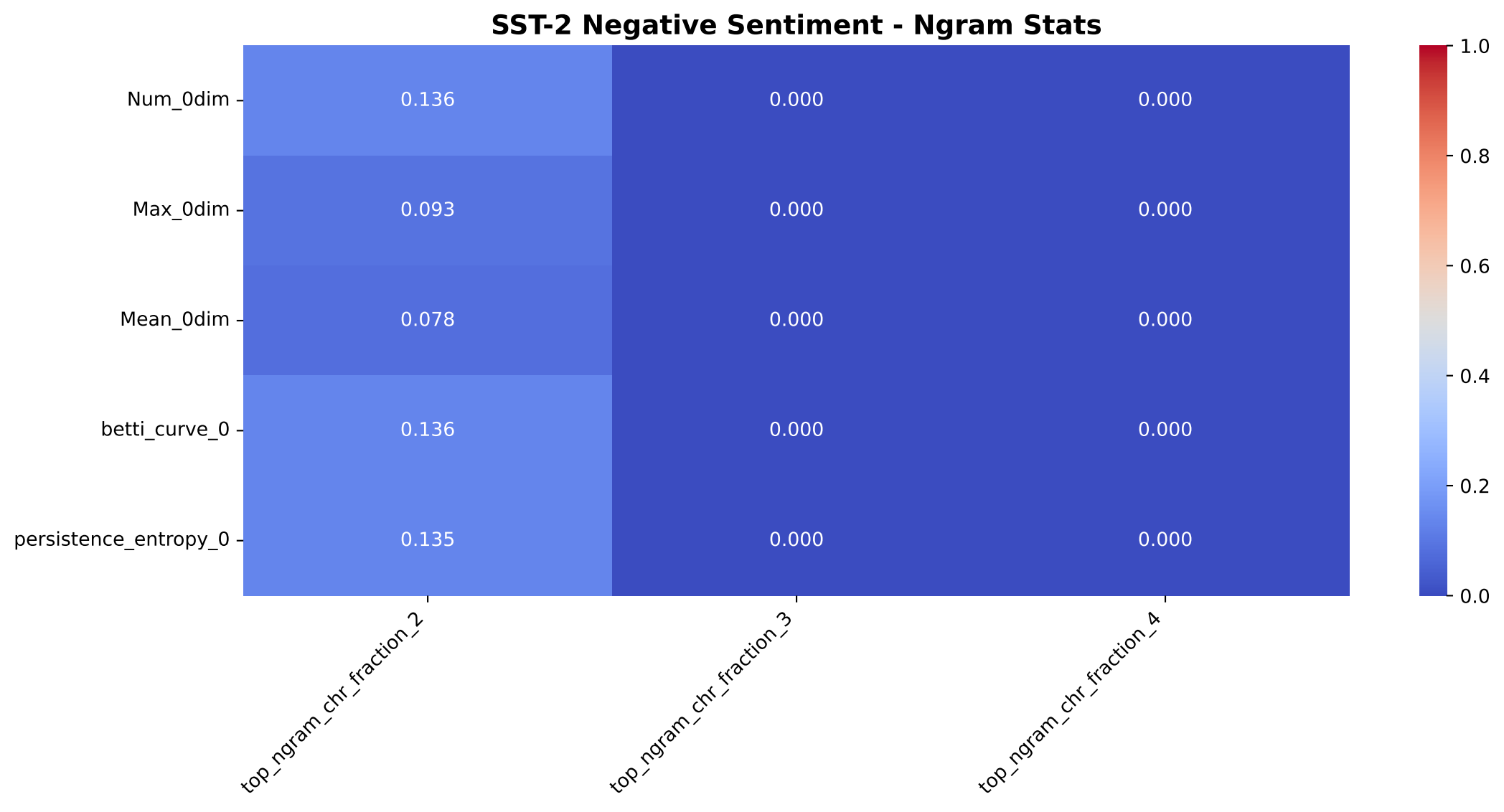

SST-2 Difference: Positive - Negative - Readability
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
-0.022 -0.071 0.000 -0.088 -0.037 -0.018 -0.048 0.016
-0.030 -0.087 6.330 -0.089 -0.057 -0.022 -0.068 -0.034
-0.022 -0.076 4.115 -0.081 -0.042 -0.048 -0.062 -0.019
-0.022 -0.071 0.000 -0.088 -0.037 -0.018 -0.048 0.016
-0.022 -0.071 0.000 -0.089 -0.038 -0.018 -0.048 0.015
flesch_reading_ease
flesch_kincaid_grade
smog
gunning_fog
automated_readability_index
coleman_liau_index
lix
rix
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

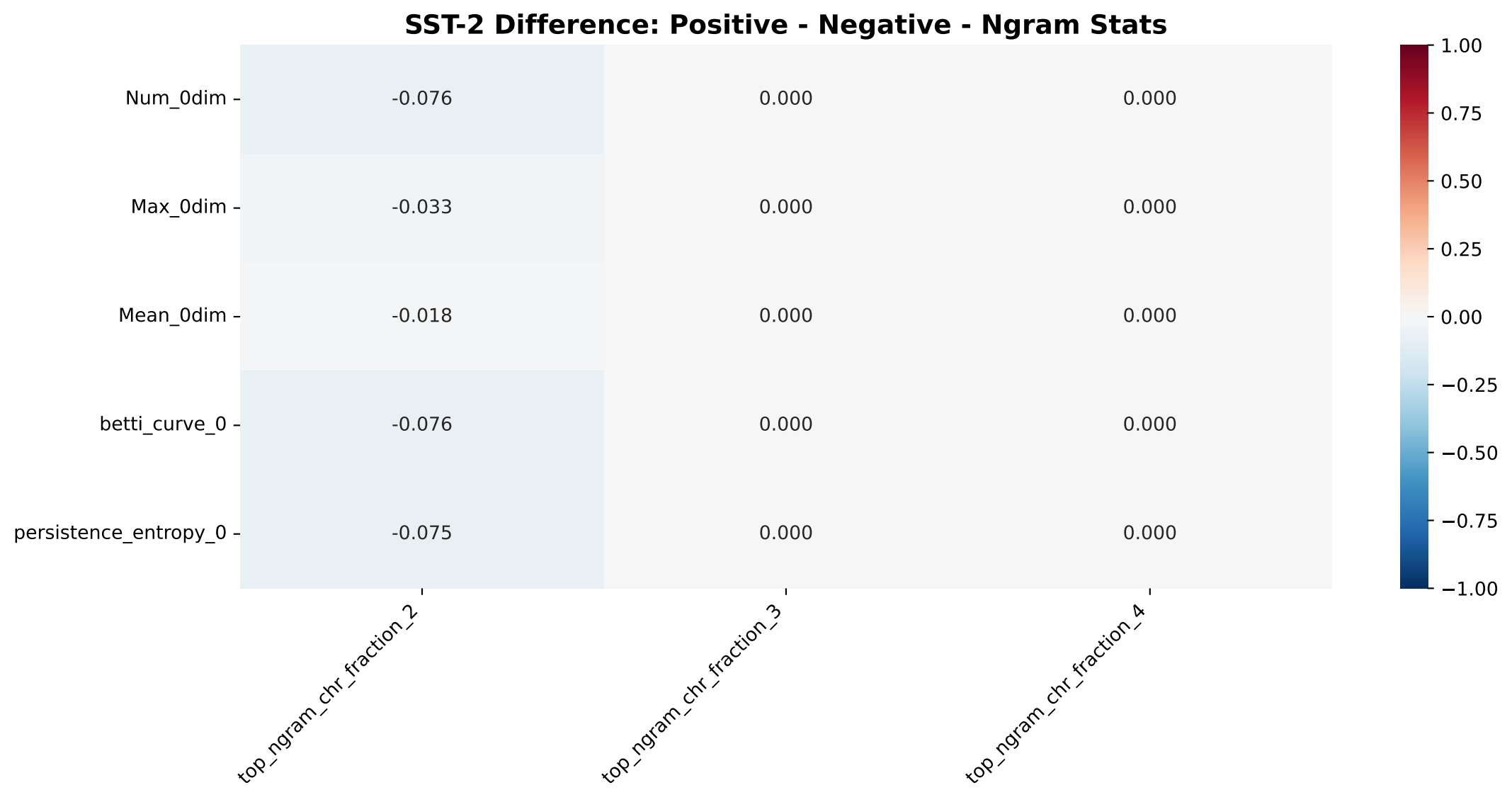

SST-2 Positive Sentiment - Quality Duplicates
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.835 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.754 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.808 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.835 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.837 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
passed_quality_check
duplicate_line_chr_fraction
duplicate_paragraph_chr_fraction
duplicate_ngram_chr_fraction_5
duplicate_ngram_chr_fraction_6
duplicate_ngram_chr_fraction_7
duplicate_ngram_chr_fraction_8
duplicate_ngram_chr_fraction_9
duplicate_ngram_chr_fraction_10
1.0
0.8
0.6
0.4
0.2
0.0

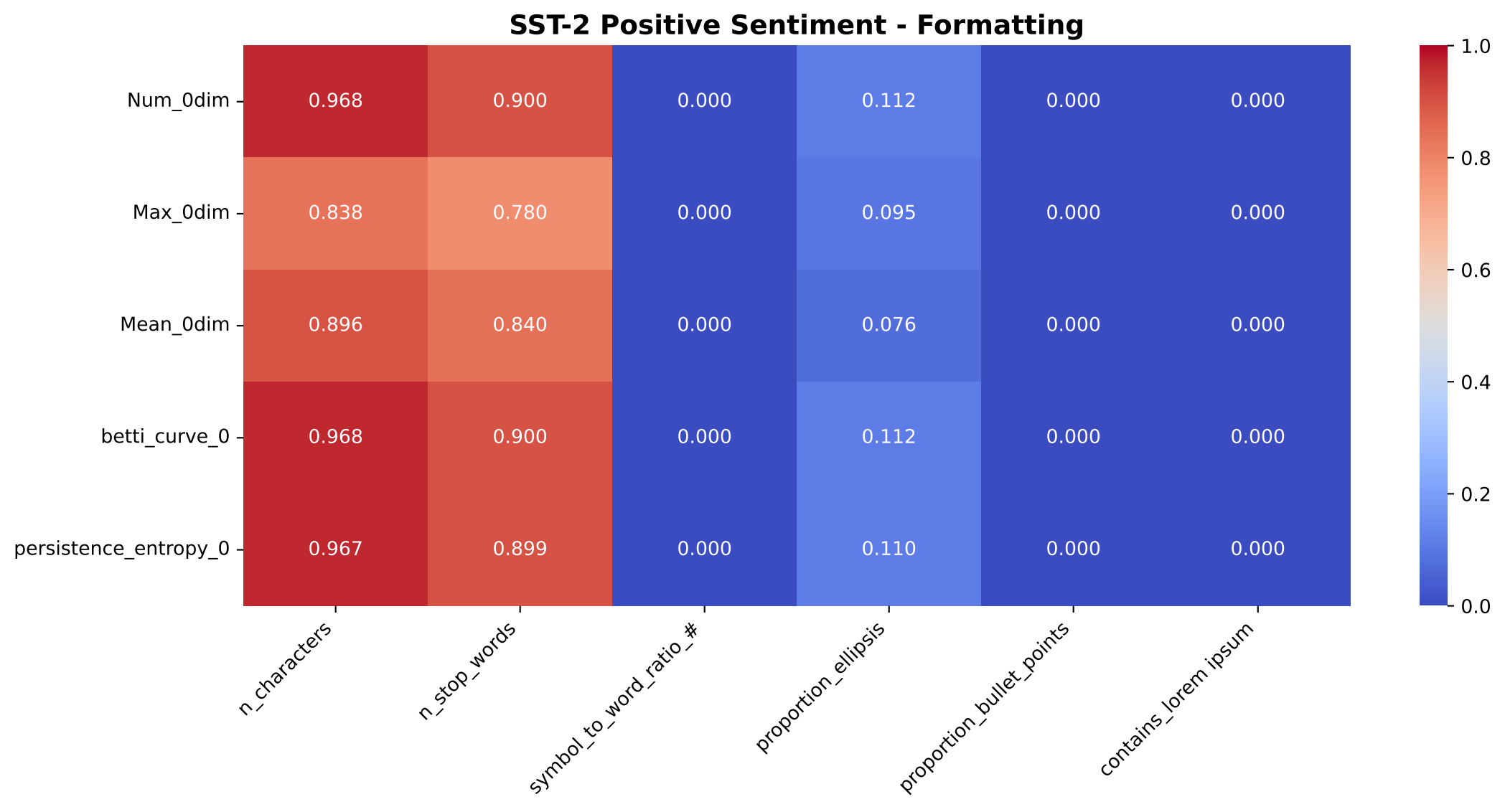

SST-2 Negative Sentiment - Quality Duplicates
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.824 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.780 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.798 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.824 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.825 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
passed_quality_check
duplicate_line_chr_fraction
duplicate_paragraph_chr_fraction
duplicate_ngram_chr_fraction_5
duplicate_ngram_chr_fraction_6
duplicate_ngram_chr_fraction_7
duplicate_ngram_chr_fraction_8
duplicate_ngram_chr_fraction_9
duplicate_ngram_chr_fraction_10
1.0
0.8
0.6
0.4
0.2
0.0

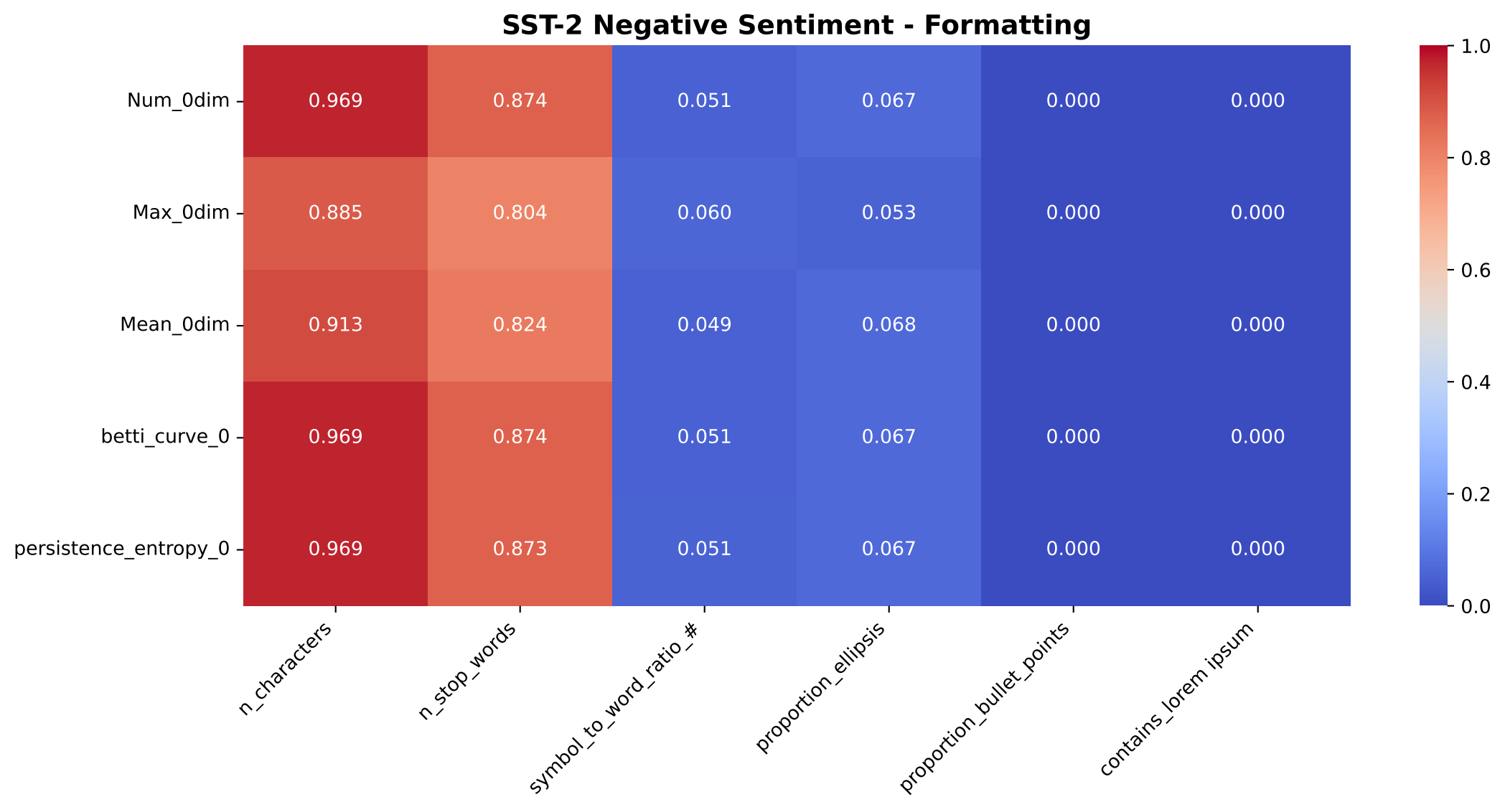

SST-2 Difference: Positive - Negative - Quality Duplicates
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.011 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
-0.027 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.010 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.011 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
0.012 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
passed_quality_check
duplicate_line_chr_fraction
duplicate_paragraph_chr_fraction
duplicate_ngram_chr_fraction_5
duplicate_ngram_chr_fraction_6
duplicate_ngram_chr_fraction_7
duplicate_ngram_chr_fraction_8
duplicate_ngram_chr_fraction_9
duplicate_ngram_chr_fraction_10
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

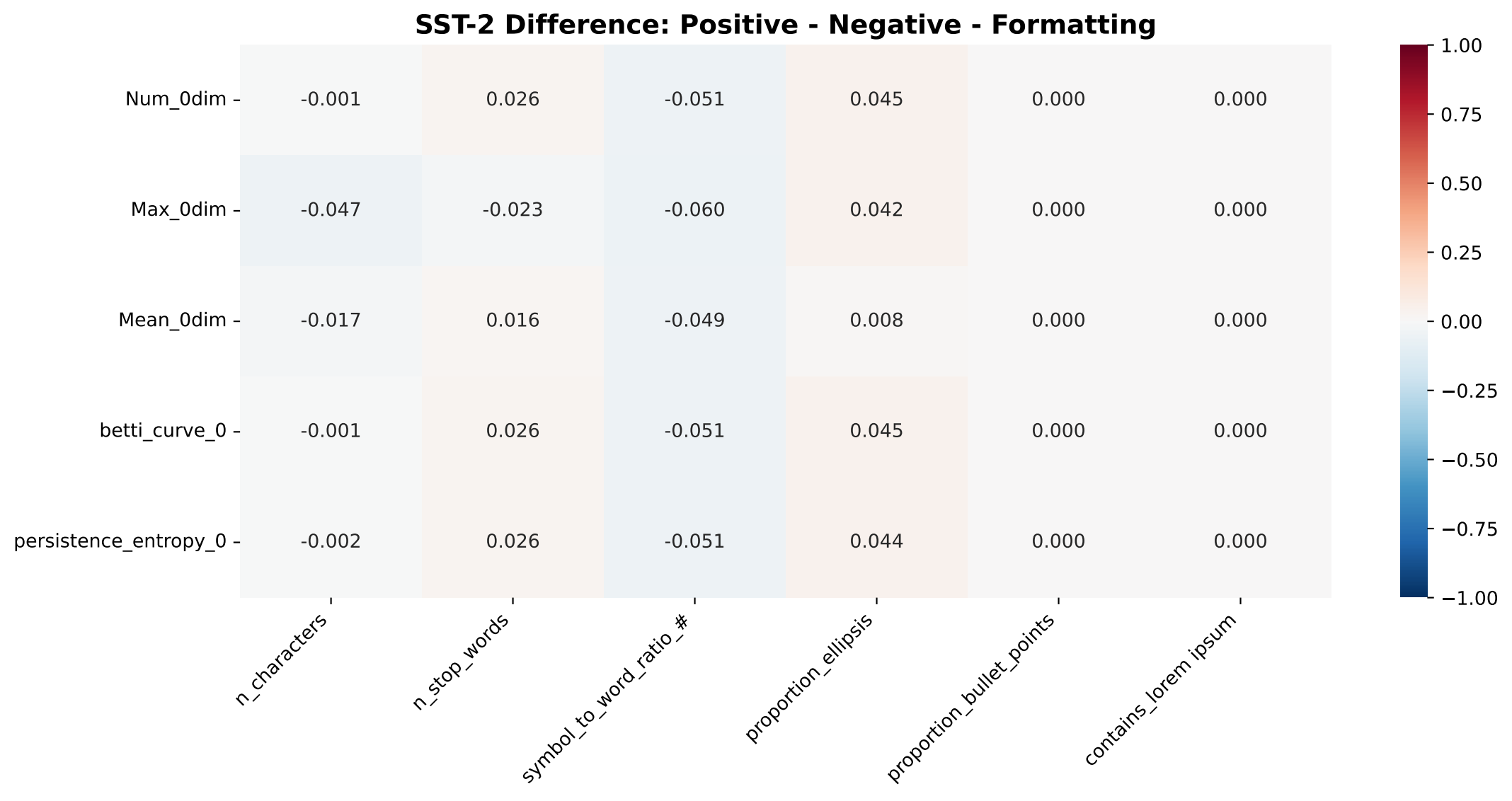

SST-2 Positive Sentiment - Ngram Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.060
0.000
0.000
0.060
0.000
0.000
0.060
0.000
0.000
0.060
0.000
0.000
0.059
0.000
0.000
top_ngram_chr_fraction_2
top_ngram_chr_fraction_3
top_ngram_chr_fraction_4
1.0
0.8
0.6
0.4
0.2
0.0


SST-2 Negative Sentiment - Ngram Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.136
0.000
0.000
0.093
0.000
0.000
0.078
0.000
0.000
0.136
0.000
0.000
0.135
0.000
0.000
top_ngram_chr_fraction_2
top_ngram_chr_fraction_3
top_ngram_chr_fraction_4
1.0
0.8
0.6
0.4
0.2
0.0


SST-2 Difference: Positive - Negative - Ngram Stats
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
-0.076
0.000
0.000
-0.033
0.000
0.000
-0.018
0.000
0.000
-0.076
0.000
0.000
-0.075
0.000
0.000
top_ngram_chr_fraction_2
top_ngram_chr_fraction_3
top_ngram_chr_fraction_4
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


SST-2 Positive Sentiment - Formatting
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.968
0.900
0.000
0.112
0.000
0.000
0.838
0.780
0.000
0.095
0.000
0.000
0.896
0.840
0.000
0.076
0.000
0.000
0.968
0.900
0.000
0.112
0.000
0.000
0.967
0.899
0.000
0.110
0.000
0.000
n_characters
n_stop_words
symbol_to_word_ratio_#
proportion_ellipsis
proportion_bullet_points
contains_lorem ipsum
1.0
0.8
0.6
0.4
0.2
0.0


SST-2 Negative Sentiment - Formatting
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
0.969 0.874 0.051 0.067 0.000 0.000
0.885 0.804 0.060 0.053 0.000 0.000
0.913 0.824 0.049 0.068 0.000 0.000
0.969 0.874 0.051 0.067 0.000 0.000
0.969 0.873 0.051 0.067 0.000 0.000
n_characters
n_stop_words
symbol_to_word_ratio_#
proportion_ellipsis
proportion_bullet_points
contains_lorem ipsum
1.0
0.8
0.6
0.4
0.2
0.0


SST-2 Difference: Positive - Negative - Formatting
Num_0dim
Max_0dim
Mean_0dim
betti_curve_0
persistence_entropy_0
-0.001
0.026
-0.051
0.045
0.000
0.000
-0.047
-0.023
-0.060
0.042
0.000
0.000
-0.017
0.016
-0.049
0.008
0.000
0.000
-0.001
0.026
-0.051
0.045
0.000
0.000
-0.002
0.026
-0.051
0.044
0.000
0.000
n_characters
n_stop_words
symbol_to_word_ratio_#
proportion_ellipsis
proportion_bullet_points
contains_lorem ipsum
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00