  - Att-Patterns-4 with an original frequency of 4 or more
  - Att-Patterns-5 with an original frequency of 5 or more

- BI-Patterns, with an original frequency of 5 or more[37]
- Unattested-Patterns
- FL-Patterns-15, FL-Patterns-30, FL-Patterns-all

Evaluating $H_0$1:

We calculated the MI and Z-sc association scores of the two words in each of the Attested-Patterns and BI-Patterns in the Diana-Araknion++ 600M token corpus. The association score was calculated based on the sentential co-occurrence of the two words. Patterns that co-occurred less than 5 times obtained a score of 0. First, we compared the obtained association with standard thresholds, representing statistical chance: 0, 0.5, and 1 for MI; 0, 1.96, and 3.29 for Z-sc. Second, we compared the average association score of the Attested Patterns with those of the BI-Patterns.

Table 3.6 shows what percentage of the Attested-Patterns in each group obtains scores higher than statistical chance. Overall, the majority of the Attested-Patterns outperform the statistical chance baseline. The results are consistent for both the measures and their thresholds, even though they measure the association in a qualitatively different manner. It is important to note that filtering out the Attested-Patterns with a frequency of 1 significantly improves the results. We believe this factor should be taken into consideration in future experiments.

Table 3.6 Association score of Attested-Patterns compared with statistical chance

	MI			Z-sc		
Patterns	**>0**	**>0.5**	**>1**	**>0**	**>1.96**	**>3.29**
Att-Patterns-5	85%	83%	80%	85%	83%	82%
Att-Patterns-4	84%	82%	79%	84%	82%	80%
Att-Patterns-3	82%	80%	77%	82%	80%	78%
Att-Patterns-2	78%	76%	72%	78%	76%	73%
Att-Patterns-all	68%	66%	62%	68%	65%	62%

As a complementary evaluation, we directly compared the association scores of the Attested-Patterns with those of the BI-Patterns. Table 3.7 shows the average association scores for the two types of patterns[38]. The Attested-Patterns have a much higher degree of association than the BI-Patterns. In the case of MI, the

[37] 5,285 of the BI-patterns coincide with Attested-Patterns.

[38] The average is calculated as a simple average of all patterns of the corresponding type.

Attested-Patterns obtain scores more than two times higher than the BI-Patterns. In the case of Z-sc, the Attested-Patterns obtain scores between 30% and 100% higher than the BI-Patterns.

Table 3.7 Average association score of Attested-Patterns and BI-patterns

Patterns	Average MI	Average Z-sc
Attested-Patterns-5	3.90	52
Attested-Patterns-4	3.86	49
Attested-Patterns-3	3.80	46
Attested-Patterns-2	3.70	42
Attested-Patterns-all	3.50	35
BI-Patterns	1.72	27

The obtained results disprove $H_0 1$ and confirm Hypothesis 1. That is, we can conclude that the Attested-Patterns are semantically coherent.

Evaluating $H_0$2.1:

We checked how many of the Unattested-Patterns were present in Diana-Araknion++. As a baseline we used the FL-Patterns. Both Unattested-Patterns and FL-Patterns are not directly obtained, but are rather a result of generalization and generation using different methodologies. For each group, we calculated the percentage of the patterns that appear once and the percentage of the patterns that appear at least five times. Table 3.8 shows the results obtained.

Table 3.8 Occurrence of Unttested-Patterns and FL-Patterns

Patterns	Occurred Once	Occurred Five Times
Unattested-Patterns	54%	24%
FL-Patterns-15	24%	9%
FL-Patterns-30	11%	4%
FL-Patterns-all	4%	0.6%

Unattested-Patterns appear much more frequently than the patterns generated by simply combining frequent lemmas. 56% of the Unattested-Patterns were observed in Diana-Araknion++. This is more than double the observance rate of the FL-Patterns-15 and five times higher than for FL-Patterns-30. 24% of the Unattested-Patterns appear in Diana-Araknion++ with a frequency of 5 or more. This is almost three times higher than FL-Patterns-15 and six times higher than FL-Patterns-30. The results of FL-Patterns-all are much lower, showing that unfiltered pattern generation is not effective. Unattested-Patterns are linguistic patterns

given that they appear in a corpus with a much higher probability than patterns generated using a simpler frequency based methodology. These results disprove $H_0$2.1.

Evaluating $H_0$2.2:

We calculated the association score (MI and Z-sc) between the lemmas in each of the Unattested-Patterns that occurred at least 5 times[39] in Diana-Araknion++. We compared the scores with the same thresholds we used when evaluating $H_0$1. Table 3.9 shows the percentage of patterns with a score higher than the statistical chance thresholds.

Table 3.9 Association scores of Unttested-Patterns

	MI			Z-sc		
Patterns	**>0**	**>0.5**	**>1**	**>0**	**>1.96**	**>3.29**
Unattested-Patterns	93%	86%	76%	93%	80%	70%

The observed degree of association is very high. Over 90% of the observed Unattested-Patterns obtained a positive association score with respect to both measures. When comparing them with the statistical chance thresholds, the obtained results are similar to those obtained by Attested-Patterns in $H_0$1. The Unattested-Patterns, when observed in a different corpus, are semantically coherent. This disproves $H_0$2.2.

In conclusion, the automated statistical evaluation of the patterns obtained by DISCOver shows that: (1) Attested-Patterns are semantically coherent, as they outperform two baselines: statistical chance thresholds and BI-Patterns. These results disprove $H_0$1.; (2) A significant percentage (56%) of the Unattested-Patterns can be found in Diana-Araknion++, which is much higher than the occurrence of FL-Patterns. These results disprove $H_0$2.1; (3) Whenever Unattested-Patterns occur in Diana-Araknion++, the statistical association between the lemmas in the patterns is much higher than the statistical chance baseline. This disproves $H_0$2.2.

As we have disproved all 3 of the null hypotheses, we can conclude that the patterns obtained by the DISCOver methodology have both properties of constructions: syntactic and semantic coherence and generalizability. Therefore they are good candidates-to-be-constructions.

We also performed a manual evaluation of the lexico-syntactic patterns. This complementary validation reinforces the results obtained in the two statistical evaluations. We prepared a dataset of 600 patterns for the manual evaluation: 300 patterns obtained by applying the DISCOver methodology (the patterns were randomly selected from all Attested and Unattested Patterns) and 300 of the FL-

[39]Calculating this score for patterns with lower frequency is unreliable due to the low-frequency bias in some of the measures.