JOURNAL ON COMMUNICATIONS

ISSN:1000-436X

Scopus®

REGISTERED

www.jocs.review

Statistical Analysis of the Similarity between Consider and Allow for in COCA and BNC

Namkil Kang[1]
(Far East University, South Korea)
Yoon Mo Yang[2]
(Far East University, South Korea)

The primary aim of this study is to examine the similarities between the expressions *consider* and *allow for* in American and British English and explore whether national variation influences their usage. In the Corpus of Contemporary American English (COCA), the genre-ranking similarity between the two expressions is 37.5% (matching rankings in 3 out of 8 genres), indicating limited overlap. In the British National Corpus (BNC), the similarity decreases to 14.28% across seven genres, further suggesting minimal genre-based alignment. Interestingly, genre-specific proximity varies: in COCA, *consider* is most dissimilar to *allow for* in academic texts and most similar in web genres; in the BNC, the greatest divergence appears in the miscellaneous genre, with the closest alignment in spoken genres. Both expressions show substantial frequency variation across genres in both corpora, with no significant national differences in their usage patterns. In COCA, *consider* appears between 7,519 and 20,137 times, while *allow for* ranges from 115 to 1,615. In the BNC, *consider* ranges from 334 to 2,938, and *allow for* from 8 to 230. Correlation and linear regression analyses further reveal strong positive relationships. In COCA, the correlation coefficient (R) is 0.8739, while in the BNC, it is 0.9186, with 84.39% of the variation in *allow for* explained by *consider*. The major finding of this study is that there is no stark difference between American and British English in seven different statistical analyses, including Chi-squared tests, which show significant differences in genre distribution across both corpora ($p < 0.000001$).

Keywords: ranking, Euclidean distance, variance, standard deviation, correlation, linear regression, Chi-squared test

1. Introduction

This paper provides an in-depth statistical analysis of the phrases *consider* and *allow for* in both American English and British English. The primary aim of this study is to examine the similarities between these two expressions across the two varieties of English and to explore whether there are any national variations in their usage. To conduct this research, we employed Python, a powerful tool for statistical analysis, and carried out seven different analyses to assess the similarities between the two phrases in American and British English. Before diving into the analysis, it is essential to clarify how the relevant data was obtained. The Corpus of Contemporary American English (COCA) [1], the British National Corpus (BNC) [2], the Hansard Corpus [3], and the Corpus of Historical American English (COHA) [4] have provided us with valuable linguistic insights. The genre frequency of *consider* and *allow for* was gathered from the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC) in August 2025. These datasets remain accessible through Google. The COCA is composed of eight genres: Blog, Web, TV/Movie, Spoken (Spok), Fiction (Fic), Magazine (Mag), Newspaper, and Academic (Acad). In contrast, the BNC consists of seven genres: Spoken (Spok), Fiction (Fic), Magazine (Mag), Newspaper, Non-academic (Non-acad), Academic (Acad), and Miscellaneous (Misc). We collected frequency data for *consider* and *allow for* from each of these genres. The analysis proceeded as follows:

1. Ranking Analysis: By analyzing the genre rankings, we examined the proximity of *consider* and *allow for* in both the COCA and the BNC. This analysis also helped determine whether any national variation exists between the two varieties.

---

[1] First author

[2] Corresponding author

2. Euclidean Distance Analysis: We calculated the Euclidean distance between the frequencies of the two phrases to assess the degree of similarity in their usage across the two corpora. Additionally, we examined any potential national differences.
3. Variance Analysis: We evaluated the distribution of the two phrases using variance to understand how their usage differs between American and British English.
4. Standard Deviation Analysis: By performing a standard deviation analysis, we explored how the frequency of *consider* and *allow for* deviate from their respective means in both varieties of English.
5. Correlation Analysis: We investigated the correlation between the two phrases to determine how closely their usage patterns align in American and British English.
6. Linear Regression Analysis: Using linear regression, we assessed how much the independent variable (*Consider*) influences the dependent variable (*Allow for*) in both varieties of English.
7. Chi-Squared Test Analysis: Finally, we conducted a Chi-squared test to determine whether there is a significant association between the two phrases as categorical variables in both varieties of English.

Through these seven statistical analyses, we gained a comprehensive understanding of the similarities and differences in the usage of *consider* and *allow for* in American and British English.

2. Data Collection

Frequency data for *consider* and *allow for* were extracted from both the COCA and the BNC in August 2025. The COCA is organized into eight distinct genres, while the BNC contains seven genres. Data for each genre from both corpora were gathered and subjected to a series of seven statistical analyses. These analyses allowed for a detailed comparison of the two expressions across different genres, facilitating an exploration of both their linguistic relationship and the potential influence of national variation on their usage patterns.

3. Results

3.1. Ranking Analysis

This section presents a ranking analysis of the phrases *consider* and *allow for* across eight genres in the Corpus of Contemporary American English (COCA) and seven genres in the British National Corpus (BNC). Ranking analysis here refers to a comparative frequency analysis that identifies which genres show a relative preference for each phrase. We begin by examining the COCA data:

Table 1 Frequency of Consider and Allow for in the COCA

Genre	All	Blog	Web	TV/M	Spok	Fic	Mag	News	Acad
Consider	110,624	21,050	19,478	6,662	8,892	5,533	16,059	12,535	20,415
Allow for	6,921	1,218	1,227	108	323	147	928	609	2,361

As shown in Table 1, *consider* occurs 110,624 times across the corpus, while *allow for* appears 6,921 times. This stark contrast suggests a strong overall preference for *consider* in American English. In other words, *consider* is significantly more frequently used than *allow for* in the COCA. These frequency counts form the basis for seven different analyses across the eight genres, as shown in Table 2.

Table 2 Ranking Analysis of Consider and Allow for in the COCA

Genre	All	Blog	Web	TV/M	Spok	Fic	Mag	News	Acad
Consider	110,624	21,050 (Rank 1)	19,478 (Rank 3)	6,662 (Rank 7)	8,892 (Rank 6)	5,533 (Rank 8)	16,059 (Rank 4)	12,535 (Rank 5)	20,415 (Rank 2)
Allow for	6,921	1,218 (Rank 3)	1,227 (Rank 2)	108 (Rank 8)	323 (Rank 6)	147 (Rank 7)	928 (Rank 4)	609 (Rank 5)	2,361 (Rank 1)

It is noteworthy that *consider* and *allow for* share the same rank in three genres: Spoken (Rank 6), Magazine (Rank 4), and Newspaper (Rank 5). This overlap indicates some degree of similarity in their distribution across these genres. However, in the remaining five genres, their rankings diverge. In quantitative terms, the ranking

similarity between the two phrases is 37.5% (3 out of 8 genres), indicating a relatively low degree of overlap in genre-specific usage. This suggests that while both *consider* and *allow for* are used across a range of genres in American English, they tend to be preferred in different contexts, reflecting differences in stylistic or functional usage.

Now attention is paid to the BNC:

Table 3 Ranking Analysis of Consider and Allow for in the BNC

Genre	All	Spok	Fic	Mag	News	Non-acad	Acad	Misc
Consider	11,456	672 (Rank 6)	709 (Rank 5)	547 (Rank 7)	806 (Rank 4)	1,877 (Rank 3)	3,609 (Rank 1)	3,236 (Rank 2)
Allow for	835	37 (Rank 5)	19 (Rank 7)	65 (Rank 4)	30 (Rank 6)	164 (Rank 3)	209 (Rank 2)	311 (Rank 1)

It is worth noting that the overall frequency of *consider* in the BNC is 11,456 tokens, compared to just 835 tokens for *allow for*. This indicates a clear preference for *consider* over *allow for* in British English. Interestingly, this pattern mirrors the findings from the COCA, where *consider* is also significantly more frequent than *allow for*. This shared preference suggests that both American and British English exhibit similar tendencies in the use of these two expressions. On the other hand, it is important to note that *consider* and *allow for* show similarities in only one genre (the non-academic genre), while differing in the other six. Specifically, the two expressions are 14.28% similar in the ranking analysis across the seven genres. This suggests that *consider* and *allow for* exhibit a relatively low degree of similarity in terms of their genre-based ranking patterns. Interestingly, there is no significant difference between American English and British English in terms of the degree of similarity between *consider* and *allow for*.

3.2. Euclidean Distance Analysis

This section is dedicated to contemplating the similarity between *consider* and *allow for* in eight genres (the COCA) and seven genres (the BNC) by employing Euclidean distance. The Euclidean distance can provide the degree of similarity between *consider* and *allow for* in each genre. We adopt the following definition of Euclidean distance:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

Now take a look at Table 4:

Table 4 Euclidean Distance in the COCA

Genre	Blog	Web	TV/M	Spoken	Fic	Mag	News	Acad
Consider (%)	19.02	17.60	6.02	8.03	5.00	14.51	11.33	18.45
Allow for (%)	17.59	17.72	1.56	4.66	2.12	13.40	8.79	34.11
Euclidean Distance	1.43	0.12	4.46	3.37	2.88	1.11	2.54	15.66

It is important to note that *consider* is furthest from *allow for* in the academic genre, with a Euclidean distance of 15.66. This suggests that *consider* and *allow for* exhibit the lowest degree of similarity in the academic genre. In contrast, *consider* is closest to *allow for* in the web genre, indicating the highest level of similarity between the two in this context. Notably, the web genre is followed closely by the magazine genre, where *consider* is the second closest to *allow for*, with a Euclidean distance of 1.11. This suggests a high degree of similarity in the

magazine genre as well. Interestingly, the blog genre ranks third, with *consider* being the third closest to *allow for* at a Euclidean distance of 1.43, indicating a moderate level of similarity. In conclusion, while *consider* is furthest from *allow for* in the academic genre, it is nearest to *allow for* in the web genre, highlighting notable genre-based variations in their usage patterns.

Now attention is paid to the BNC:

Table 5 Euclidean Distance in the BNC

Genre	Spoken	Fic	Mag	News	Non-acad	Acad	Misc
Consider (%)	5.86	6.18	4.77	7.03	16.38	31.50	28.24
Allow for (%)	4.43	2.27	7.78	3.59	19.64	25.02	37.24
Euclidean Distance	1.43	3.91	3.01	3.44	3.26	6.48	9.00

It is important to note that *consider* is furthest from *allow for* in the miscellaneous genre, with a Euclidean distance of 9. This suggests that the two expressions exhibit the lowest degree of similarity in this genre. In contrast, *consider* is closest to *allow for* in the spoken genre, indicating the highest level of similarity between the two in this context. The spoken genre is followed by the magazine genre, where *consider* ranks as the second closest to *allow for*, with a Euclidean distance of 3.01, suggesting a high degree of similarity in the magazine genre as well. Additionally, *consider* is the third closest to *allow for* in the non-academic genre, with a Euclidean distance of 3.26, implying that they show the third highest similarity in this genre. In conclusion, while *consider* is furthest from *allow for* in the miscellaneous genre, it is closest to *allow for* in the spoken genre, highlighting distinct genre-based patterns in their usage.

3.3. Variance

In what follows, we examine the extent to which the frequency of *consider* and *allow for* varies across different genres in two major corpora: COCA (Corpus of Contemporary American English) and the BNC (British National Corpus). This analysis focuses on the variance, a statistical measure that reflects how spread out the frequency values are from their respective means. Let us begin with Table 6, which presents the mean and variance for each expression in COCA:

Table 6 Variance (COCA)

	Mean	Variance
Consider	13,828	39,805,748
Allow for	865.125	563,644

Variance, as a statistical concept, indicates the degree of dispersion in a dataset—that is, how far individual data points deviate from the mean. In the case of *consider*, the exceptionally high variance of nearly 40 million reveals that its frequency varies substantially across genres in American English. For example, in some genres, *consider* may appear over 20,000 times, while in others, its frequency drops to around 5,500 occurrences. This wide range indicates that the use of *consider* is not evenly distributed across COCA's genre divisions and suggests that it is highly sensitive to contextual factors, such as genre or domain (e.g., academic, fiction, newspaper, spoken discourse, etc.). Turning to *allow for*, we see a mean frequency of approximately 865 occurrences, but a variance of 563,644, which—relative to the mean—is also notably high. This indicates that the phrase *allow for* exhibits significant variability in its usage across genres in American English. For instance, one genre may feature as few as 108 occurrences, while another records as many as 2,361. Such disparity highlights the context-dependent nature of *allow for*, which may be more prevalent in technical or instructional registers than in narrative or conversational texts. Now, let us consider the same expressions in the British National Corpus, as presented in Table 7:

Table 7 Variance (BNC)

	Mean	Variance
Consider	1,636	1,695,708
Allow for	119	12,424

Again, we observe high variance values for both expressions, despite the lower mean frequencies compared to COCA. The variance for *consider* in the BNC is approximately 1.7 million, suggesting that its distribution across British English genres is likewise uneven. Some genres appear to use the word *consider* much more frequently than others, reinforcing the idea that its usage is heavily genre-specific. *Allow for* is a less frequently used phrase overall in British English. The variance, although small in absolute terms, is large relative to the mean.

3.4. Standard Deviation

The standard deviation provides valuable insight into how much the frequency of a word or phrase deviates from the average (mean) across different genres. Let us consider Table 8:

Table 8 Standard Deviation (COCA)

	Mean	Variance	Standard Deviation
Consider	13,828	39,805,748	6,309
Allow for	865.125	563,644	750

In the case of *consider* in COCA, the standard deviation is approximately 6,309, which indicates a substantial level of variability in its usage across genres. Statistically, this suggests that the frequency of *consider* typically falls within the range of approximately 7,519 to 20,137 (i.e., one standard deviation below and above the mean). In other words, although the average frequency is 13,828, actual frequencies across genres tend to fluctuate by about 6,309 in either direction. This wide dispersion reinforces the earlier observation that *consider* is a high-frequency word in American English whose usage is heavily genre-dependent. In contrast, the phrase *allow for* has a much lower mean frequency of 865.125 and a standard deviation of approximately 750. This, too, suggests considerable variability in its distribution, albeit on a smaller scale due to its overall lower frequency. The data indicates that *allow for* typically appears with a frequency between 115 and 1,615 in different genres. The large standard deviation relative to its mean further implies that this phrase is highly context-sensitive, with significantly more occurrences in some genres than others, particularly in technical, instructional, or formal domains. Now take a look at Table 9.

Table 9 Standard Deviation (BNC)

	Mean	Variance	Standard Deviation
Consider	1,636	1,695,708	1302
Allow for	119	12,424	111

Turning now to the British National Corpus (BNC), we observe similar trends, though the absolute frequencies are generally lower than in COCA. The standard deviation for *consider* in the BNC is approximately 1,302, which suggests that its frequency commonly ranges from 334 to 2,938 across different genres. This again reflects a notable degree of variability in its use, implying that *consider* plays an important role in some genres (likely academic or formal writing) while being less common in others (such as fiction or spoken English). Interestingly, the standard deviation for *allow for* in the BNC is approximately 111, which indicates that its frequency commonly ranges from 8 to 230. The standard deviation values for both corpora confirm that *consider* is a relatively common word in both American and British English, though its frequency is subject to considerable variation across genres. Meanwhile, *allow for*—though used less frequently overall—shows even greater dispersion, indicating that it is much more context- and genre-specific.

3.5. Correlation Analysis

In this section, we aim to investigate the correlation between the two phrases to determine how closely their usage patterns align in American and British English. Now take a look at Table 10:

Table 10 Pearson Correlation Coefficient: COCA

		Allow for
Consider	*r*	0.8739
	p	0.0045

	N	8

A Pearson's *r* of 0.8739 indicates a very strong positive correlation. This means that the frequencies of *consider* and *allow for* across genres in COCA tend to rise and fall together. When one is frequent in a genre, the other is also likely to be frequent.

Table 11 Scale of Interpretation for Pearson's *r*:

R value	Strength
0.00-0.19	Very weak
0.20-0.39	Weak
0.40-0.59	Moderate
0.60-0.79	Strong
0.80-1.00	Very strong

So, 0.8739 is very strong. A p-value of 0.0045 is well below the standard threshold of 0.05, indicating the correlation is statistically significant. In simpler terms, there is strong evidence that the observed correlation is real, not just a random coincidence. The correlation is based on 8 paired data points (i.e., 8 genre frequencies). While the sample is small, the high *r* and low *p*-value suggest that the correlation is still robust. However, with small N, caution is needed in generalizing the result beyond these specific genres. Now let us consider the BNC:

Table 12 Pearson Correlation Coefficient: BNC

		Allow for
Consider	*r*	0.9186
	p	0.0035
	N	7

An *r*-value of 0.9186 represents an extremely strong positive correlation. This suggests that across the 7 genres analyzed, as the frequency of *consider* increases, the frequency of *allow for* also increases in a nearly linear fashion. So, 0.9186 is close to perfect correlation. The p-value of 0.0035 indicates that the correlation is highly statistically significant. Therefore, we can confidently conclude that the correlation is not accidental. Although N = 7 is a small sample, the extremely high *r*-value and the statistically significant *p*-value suggest that the relationship is strong and reliable—within this dataset. However, generalization beyond these 7 data points should still be done with caution.

3.6. Linear Regression Analysis

In this section, we aim to evaluate the extent to which the frequencies of *consider* influence the frequencies of *allow for*.

Table 13 Model Summary: COCA

Model	R	R^2	Adjusted R^2	Std. Error
1	0.8739112615737038	0.7637208931053427	0.7243410419562332	1.6676593029254039

The model shows a strong positive correlation (R = 0.8739), which indicates that the independent variable (consider) has a solid relationship with the dependent variable (allow for). The model explains 76.37% of the variance in the dependent variable ($R^2 = 0.7637$), which is a good fit for most practical purposes. This means that the model is able to predict the dependent variable quite well based on the independent variable. After accounting for the number of predictors in the model, 72.43% of the variance is explained, which suggests that the model is not overfitting and has a strong predictive power that holds even with multiple predictors. The standard error (1.6677) shows that, on average, the predicted values differ from the observed values by this amount. While this

is reasonable, there may still be opportunities to refine the model for even more precise predictions. Now take a look at Table 14:

Table 14 Coefficients (COCA)

Model	B	Std. Error	Beta	*t*	*p*
1 (Constant)	-572.8668060794093	355.0282083956936	-68.16844118425783	-1.613581097310 7965	0.1577452117801 2615
Consider	0.10399130793169001	0.02361384671319273	0.8739112615737704	4.403827516742 264	0.0045495431982 6914

The intercept term is not statistically significant (p-value = 0.1577). This suggests that the baseline value of the dependent variable (when consider is 0) is not reliably different from zero. *Consider* has a positive and statistically significant effect on the dependent variable, with a coefficient of 0.1040. This indicates that as the frequency of *consider* increases, the dependent variable (allow for) also increases, and the effect is statistically significant (p-value = 0.0045). The Beta coefficient for *consider* is 0.8739, suggesting that *consider* is a strong predictor of the dependent variable. The t-value of 4.4038 supports this significance, indicating that the effect of *consider* is unlikely to be due to random chance. Now attention is paid to the BNC:

Table 15 Model Summary: BNC

Model	R	R^2	Adjusted R^2	Std. Error
1	0.9186225387721477	0.8438673687401861	0.8126408424882233	0.015126300190955316

The model shows a very strong positive relationship between *consider* and *allow for* (R = 0.9186). 84.39% of the variation in the dependent variable (allow for) is explained by the independent variable (consider), which is a high explanatory power (R^2 = 0.8439). The Adjusted R^2 = 0.8126 further confirms the strength of the model, even after adjusting for the number of predictors. The small standard error (0.0151) indicates that the model's predictions are very accurate, suggesting that the model fits the data well. Overall, this model appears to be a strong and reliable predictor of the dependent variable (allow for), with *consider* explaining a significant portion of the variation in the outcome. Now let us turn to Table 16:

Table 16 Coefficients (BNC)

Model	B	Std. Error	Beta	*t*	*p*
1 (Constant)	-9.403734153839544	30.747087399117575	-109.85743766720745	-0.305841461721 97846	0.7720496083072 428
Consider	0.07863356660936426	0.01512630019095 5316	0.0067309885581 23103	5.198466618848 589	0.0034714703609 916983

The intercept term is not statistically significant (p-value = 0.7721), meaning there is insufficient evidence to suggest that the baseline value of the dependent variable (allow for) is different from 0 when *consider* is 0. *Consider* has a positive and statistically significant effect on *allow for* with a coefficient of 0.0786 (p-value = 0.0035). This means that as the frequency of *consider* increases, the frequency of *allow for* is expected to increase by 0.0786 units, and this relationship is statistically significant. Despite the positive relationship, the Beta for *consider* (0.0067) is very small, suggesting that while *consider* is a significant predictor, its effect on *allow for* is weak in standardized terms.

3.7. Chi-Squared Test Analysis

In what follows, we aim to evaluate whether there is a significant association between the *consider* frequencies and the *allow for* frequencies. Let us take a look at Table 17:

Table 17 Chi-Squared Test: COCA

Chi-Squared Statistic	1305.4179
p-value	0.000000

The Chi-squared test was conducted to examine whether there is a statistically significant difference in the distribution of the expressions *consider* and *allow for* across the eight genres of the Corpus of Contemporary American English (COCA). The test yielded a Chi-squared statistic of 1305.42 with 7 degrees of freedom, and the associated p-value was less than 0.000001. Given the extremely low p-value, we can confidently reject the null hypothesis, which assumes that the two expressions are distributed similarly across the different genres. This result indicates that the observed differences in frequency between *consider* and *allow for* are not due to random variation but reflect meaningful differences in usage patterns. In practical terms, this suggests that *consider* and *allow for* are used with differing frequency across various registers or genres represented in the COCA, such as spoken language, fiction, newspapers, academic writing, and so on. These differences may point to stylistic, contextual, or functional preferences associated with each expression, and they warrant further qualitative investigation to understand the specific contexts in which each form tends to appear more prominently. Now let us turn to the BNC:

Table 18 Chi-Squared Test: BNC

Chi-Squared Statistic	87.9257
p-value	0.000000

A Chi-squared test was conducted to determine whether there is a statistically significant difference in the distribution of the expressions *consider* and *allow for* across the seven genres of the British National Corpus (BNC). The test produced a Chi-squared statistic of 87.93 with 6 degrees of freedom and a p-value of less than 0.000001. Since the p-value is extremely small ($p < 0.000001$), we reject the null hypothesis, which posits that the two expressions are distributed similarly across the corpus sections. The results indicate a highly significant difference in how *consider* and *allow for* are used in different genres or registers represented in the BNC. This suggests that the two expressions are not interchangeable in practice and tend to occur in different contexts or types of texts. For example, one expression might be more prevalent in formal writing, while the other is more common in spoken or informal language. These findings highlight functional and stylistic variation in usage, and they point to the importance of considering context when analyzing or teaching such expressions.

4. Discussion

The primary aim of this study is to examine the similarities between the expressions *consider* and *allow for* across two varieties of English—American and British—and to explore whether any significant national variation exists in their usage patterns. This analysis is crucial for understanding not only how these expressions behave within different genres but also whether their distribution varies based on national linguistic norms. In the Corpus of Contemporary American English (COCA), the genre-ranking similarity between *consider* and *allow for* is 37.5%, meaning they share the same ranking in 3 out of 8 genres. This suggests a relatively low degree of overlap in their genre-specific usage in American English. In contrast, the British National Corpus (BNC) shows an even lower similarity of 14.28% across the seven genres analyzed, indicating that these two expressions have limited similarity in their genre-based distribution in British English as well. Interestingly, the proximity between *consider* and *allow for* varies across different genres within each corpus. In COCA, for example, the two expressions are most dissimilar in the academic genre, yet they show the closest alignment in the web genre. This suggests that genre-specific factors play a significant role in the usage of these expressions in American English. In the BNC, *consider* is furthest from *allow for* in the miscellaneous genre, but their relationship is closest in the spoken genre, again demonstrating notable genre-based variations in British English. These differences suggest that while both expressions exhibit some commonality across varieties of English, their genre-based usage patterns differ in significant ways. A closer look at frequency data reveals substantial variation in the usage of *consider* and *allow*

for across different genres. In COCA, *consider* typically appears between 7,519 and 20,137 times, while *allow for* ranges from 115 to 1,615. In the BNC, *consider* has a standard deviation of approximately 1,302, with frequencies ranging from 334 to 2,938, while *allow for* exhibits a standard deviation of around 111, typically appearing between 8 and 230 times. This indicates that *consider* has a broader frequency range in both corpora compared to *allow for*, yet both expressions show significant variation across genres, suggesting that genre-specific context strongly influences their usage in both American and British English. The study also includes a detailed statistical analysis to better understand the relationship between *consider* and *allow for* in both corpora. Correlation analysis reveals a notable pattern: in COCA, the frequencies of *consider* and *allow for* tend to increase and decrease together across genres. This suggests a positive, albeit moderate, relationship between the two expressions in American English. In the BNC, the relationship is even stronger, with the frequency of *allow for* rising almost linearly with that of *consider*. Linear regression analysis confirms these findings, showing a strong positive correlation between the two expressions in both corpora. In COCA, the correlation coefficient (R) is 0.8739, while in the BNC, it is even higher at 0.9186. This means that a significant portion of the variation in the frequency of *allow for* can be explained by the frequency of *consider*. Specifically, in the BNC, 84.39% of the variation in *allow for* is accounted for by *consider*. Finally, to test whether the distribution of these expressions differs significantly across genres, Chi-squared tests were conducted. In COCA, the Chi-squared statistic was 1305.42 with 7 degrees of freedom and a p-value of less than 0.000001, allowing us to confidently reject the null hypothesis that assumes similar distribution patterns. Similarly, in the BNC, the Chi-squared statistic was 87.93 with 6 degrees of freedom and a p-value also below 0.000001, further supporting the conclusion that the two expressions are distributed differently across genres in both corpora. The major finding of this study is that there is no stark difference between American and British English in terms of the genre-specific distribution of *consider* and *allow for*. Despite some variation in genre patterns and frequency ranges, both expressions exhibit similar trends in their genre-based behavior across the two varieties of English, suggesting that genre-specific factors rather than national differences are the primary influence on their usage. This finding is supported by multiple statistical analyses, including correlation, linear regression, and Chi-squared tests, which provide strong evidence that while *consider* and *allow for* may differ slightly across genres, their overall usage patterns in American and British English are largely consistent. For more information on corpus linguistics, refer to Aarts & Granger (1993), Baker, Hardie & McEnery (2006), Barlow& Kuperman (2008), Biber (1993), Biber, Conrad & Reppen (1998), Channell (2000), Firth (1957), Gries (2013), Hunston & Francis (2000), and Kilgarriff & Grefenstette (2003) [5-14].

5. Conclusion

The primary aim of this study is to examine the similarities between the expressions *consider* and *allow for* across two varieties of English—American and British—and to investigate whether any national variation exists in their usage. Notably, in the Corpus of Contemporary American English (COCA), the genre-ranking similarity between the two expressions is 37.5% (i.e., they share the same ranking in 3 out of 8 genres), indicating a relatively low degree of overlap in genre-specific usage. In the British National Corpus (BNC), this similarity drops to 14.28% across the seven genres analyzed, further suggesting that *consider* and *allow for* exhibit limited similarity in terms of their genre-based distribution. Interestingly, genre-based patterns of proximity between the two expressions vary within each corpus. In COCA, *consider* is most dissimilar to *allow for* in the academic genre but most similar in the web genre. In the BNC, *consider* is furthest from *allow for* in the miscellaneous genre and closest in the spoken genre, again highlighting distinct genre-specific usage trends. It is also notable that the frequency of *consider* shows substantial variation across genres in American English, while *allow for* exhibits significant variability as well. In the BNC, although both expressions have lower mean frequencies than in COCA, they display high variance values, suggesting consistent variability across genres in British English as well. This variability across both corpora indicates that there is no significant national difference in how the two expressions behave with respect to genre-based variation. In terms of frequency ranges, *consider* in COCA typically appears between approximately 7,519 and 20,137 instances across different genres, whereas *allow for* ranges between 115 and 1,615. In the BNC, the standard deviation for *consider* is approximately 1,302, with frequencies commonly ranging from 334 to 2,938. For *allow for*, the standard deviation is around 111, corresponding to a typical frequency range of 8 to 230 across genres. Correlation analysis further reveals notable patterns. In COCA, the frequencies of *consider* and *allow for* tend to increase and decrease together across genres. Similarly, in the BNC, the frequency of *allow for* rises almost linearly with that of *consider*. Linear regression analysis confirms a strong positive relationship between the two expressions in both corpora: in COCA, the correlation coefficient (R) is 0.8739, while in the BNC, the coefficient is even stronger at 0.9186. In the BNC model, 84.39% of the variation in the dependent variable (*allow for*) is explained by the independent variable (*consider*). Finally, Chi-squared tests were conducted to assess whether the distribution of the two expressions differs significantly across genres. In COCA, the test yielded a Chi-squared statistic of 1305.42 with 7 degrees of freedom and a p-value of less than 0.000001. This extremely low p-value allows us to confidently reject the null hypothesis, which assumes similar

distribution patterns. Likewise, in the BNC, the Chi-squared statistic was 87.93 with 6 degrees of freedom, again yielding a p-value below 0.000001. This provides strong evidence that *consider* and *allow for* are distributed differently across genres in both corpora.

References

[1] Aarts, B., & Granger, S. (1993). *Corpus-based approaches to grammar*. Benjamins Publishing.
[2] Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh University Press.
[3] Barlow, M., & Kuperman, V. (2008). *Corpus linguistics: A guide to the theory and practice*. Palgrave Macmillan.
[4] Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257.
[5] Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
[6] British National Corpus (BNC). Online https://corpus.byu.edu/bnc
[7] Channell, J. (2000). Corpus linguistics: A resource book for students. Routledge. Cambridge University Press.
[8] Corpus of Contemporary American English (COCA). Online https://corpus.byu.edu/coca.
[9] Corpus of Historical American English (COHA). Online. https://corpus.byu.edu/coha.
[10] Firth, J. R. (1957). *Papers in linguistics* 1934-1951. Oxford University Press.
[11] Gries, S. T. (2013). *Statistics for linguists: A step-by-step guide*. Mouton de Gruyter.
[12] Hansard Corpus (HC). Online https://english-corpora.org /hansard/
[13] Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins Publishing.
[14] Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on corpus-based approaches to lexical semantics. *Computational Linguistics*, 29(4), 493-500.