PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO PUC-SP

Maria Carolina Zuppardi

Collocation dimensions in academic English

Doutorado em Linguística Aplicada e Estudos da Linguagem

São Paulo

2020

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO PUC-SP

Maria Carolina Zuppardi

Collocation dimensions in academic English

Doutorado em Linguística Aplicada e Estudos da Linguagem

Tese apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para a obtenção do título de Doutora em Linguística Aplicada e Estudos da Linguagem, sob a orientação do Prof. Dr. Antônio Paulo Berber Sardinha.

São Paulo

2020

Banca Examinadora

_____

_____

_____

_____

_____

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.
Número do Processo: 88887.151033/2017-00

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001.
Process Number: 88887.151033/2017-00

AGRADECIMENTOS

Primeiramente, agradeço profundamente ao meu orientador e mentor Prof. Dr. Tony Berber Sardinha por me acompanhar, me ensinar e me guiar no mundo da pesquisa, por ser um ser humano incrível, humilde e profissional. Há oito anos, você permitiu que um mundo novo se abrisse e espero continuar essa caminhada, retribuindo toda dedicação e contribuindo para o conhecimento.

À Comissão Fulbright Brasil pela concessão da bolsa DDRA. Foram nove meses de doutorado sanduíche em que a minha pesquisa se transformou, deu um salto gigantesco e me proporcionou a oportunidade de fazer cursos, ir a congressos, trocar experiências e aprender com pesquisadores do mundo todo.

Aos professores Randi Reppen e Douglas Biber por me acolherem, por todo o apoio, ensinamento, incentivo, questionamento e por possibilitar a transformação da minha pesquisa em algo tão rico. Ao professor Jesse Egbert pela orientação, pelo incentivo a me enveredar pelo mundo da programação e por todas as contribuições a esta pesquisa. Ao professor Joe Collentine pela paciência e parceria em me ensinar e escrever comigo os scripts que tornaram esta pesquisa possível e, principalmente, pela amizade, sua e da Karina.

À Deise Dutra por me acolher em sua família, pelo apoio, conversas, discussões sobre a pesquisa e ensinamentos. À Larissa Goulart, minha melhor companheira em Flagstaff e amiga nessa e nas próximas caminhadas.

Ao GELC, grupo de pesquisa que está ampliando e disseminando a Linguística de Corpus no Brasil, com pesquisadores excepcionais, que se apoiam e crescem juntos, levando suas pesquisas para o mundo. À Marcia Veirano Pinto, por todo apoio, trocas de ideias e dedicação, pelo exemplo do que é ser pesquisadora.

À grande colega e amiga Cristina Mayer, que me “apresentou” ao GELC, me acolheu e se tornou uma grande amiga. Obrigada por todas as conversas, pelo apoio constante, pelos papos, companheirismo e troca de ideias e pela dedicação.

Ao LAEL e, em especial, à Maria Lúcia Reis, por todo auxílio nas questões burocráticas e companheirismo e ajuda nas comissões.

Ao Rodrigo, que sempre me incentivou a crescer, buscar mais e lutar pelo que eu quero.

Aos meus pais e família, que sempre me incentivaram e torcem por mim.

RESUMO

Esta pesquisa teve como objetivo principal identificar e caracterizar redes colocacionais ampliadas, ou “dimensões de colocação”, conforme são representadas na escrita acadêmica em língua inglesa, por meio de uma abordagem empírica *bottom-up*, de acordo com Berber Sardinha (2017). A existência de redes colocacionais está documentada na literatura, assim, partimos da premissa de que redes colocacionais possam existir para colocados compartilhados por nódulos, o que seria evidenciado em uma dimensão de colocação. Portanto, esta pesquisa busca ampliar o conceito de redes colocacionais ao identificar os principais grupos de padrões colocacionais compartilhados por nódulos, ou “dimensões de colocação”, na escrita acadêmica. O *Corpus of Academic Texts (CAT)* foi compilado para esta pesquisa e é composto por livros-texto e artigos científicos e está dividido em cinco *subcorpora*, um por área do conhecimento, de acordo com a *National Science Foundation* (NSF): Ciências Biológicas; Ciências da Computação e Informação e Engenharia; Engenharia; Ciências Físicas e Matemáticas; e Ciências Sociais, Comportamentais e Econômicas. Para atingir o objetivo de pesquisa, procedimentos de estatística multivariada foram empregados, tais como análise fatorial exploratória e análise de agrupamentos. A análise fatorial exploratória identificou 11 dimensões de colocação, representando dimensões de padrões colocacionais ampliados, com um total de 10.474 colocações distintas, a saber: (1) Finanças e economia; (2) Classificação; (3) Desenvolvimento cultural e social; (4) Exercendo influência e moldando resultados; (5) Métodos de pesquisa empírica; (6) Sociedade local e global; (7) Relatando pesquisa; (8) Indivíduos na sociedade; (9) Aquisição de conhecimento e evolução de conceitos; (10) Bem-estar social; (11) Marcação de tempo e texto. A análise de agrupamentos revelou cinco agrupamentos, refletindo grupos semânticos de nódulos.

Palavras-chave: Dimensão de colocação; Colocação; escrita acadêmica; Linguística de Corpus.

ABSTRACT

This study set out to identify and characterize extended collocational networks, or “collocational dimensions”, as they are represented in written academic English, through a bottom-up approach, following Berber Sardinha (2017). It has been established in the literature that collocational networks exist for individual node words hence collocational networks could exist for collocates across nodes, which would be evidenced in a dimension of collocation. Thus, this study set out to extend the concept of collocational networks by identifying the major sets of collocational patterns across nodes, or “collocational dimensions”, in academic writing. The Corpus of Academic Texts (CAT) was designed and compiled for this study. It comprises textbooks and journal articles, divided into five subcorpora, one per discipline, according to research discipline areas of the National Science Foundation (NSF): Biological Sciences; Computer and Information Science and Engineering; Engineering; Mathematical and Physical Sciences; and Social, Behavioral and Economic Sciences. To accomplish this goal, multivariate statistical procedures, such as exploratory factor analysis and cluster analysis were employed. The exploratory factor analysis led to the identification of 11 collocational dimensions, i.e., dimensions of extended collocational patterns, comprising a total of 10,474 unique collocations, namely: (1) Finance and Economics; (2) Classification; (3) Social and Cultural Development; (4) Exerting Influence and Shaping Outcomes; (5) Empirical Research Methods; (6) Local and Global Society; (7) Reporting Research; (8) Individuals in Society; (9) Gaining Knowledge and Developing Concepts; (10) Social Welfare; (11) Marking Time and Sectioning. The cluster analysis revealed five clusters, which reflected semantic groupings of nodes.

Keywords: Collocational Dimension; Collocation; Academic writing; Corpus Linguistics.

SUMÁRIO

Introduction 13
Chapter 1. Literature Review 19
1.1. Collocations 19
1.2. Lexical and Collocational Networks 20
1.3. Visualizing Collocational Networks 27
1.4. Dimensional Collocational Analysis 31
1.4.1. Theoretical Considerations 32
1.4.2. Methodological Considerations 36
Chapter 2. Methodology 42
2.1. Corpus Design and Compilation 42
2.2. Corpus Tagging 45
2.3. Collocation Identification and Extraction 48
2.3.1. Identification of Node and Collocate Pairs 49
2.3.2. Calculation of Strength of Association 53
2.3.3. Collocation Extraction 56
2.4. Exploratory Factor Analysis 58
2.4.1. Factorability of Data 60
2.4.2. Exploratory Factor Analysis 64
2.4.3. Factor Scores 76
2.4.4. Interpretation and naming of factors 79
2.5. Cluster Analysis 80
2.5.1. Assumptions 81
2.5.2. Initial Hierarchical Cluster Analysis (HCA) 83
2.5.3. Hierarchical Cluster Analysis (HCA) – Multiple clusters 88
Chapter 3. Results 94
3.1. Collocational Dimensions of Written Academic English 94
3.1.1. Collocational Dimension 1: Finance and Economics 95
3.1.2. Collocational Dimension 2: Classification 100
3.1.3. Collocational Dimension 3: Social and Cultural Development 104
3.1.4. Collocational Dimension 4: Exerting Influence and Shaping Outcomes 109
3.1.5. Collocational Dimension 5: Empirical Research Methods 113
3.1.6. Collocational Dimension 6: Local and Global Society 117
3.1.7. Collocational Dimension 7: Reporting Research 121
3.1.8. Collocational Dimension 8: Individuals in Society 125
3.1.9. Collocational Dimension 9: Gaining Knowledge and Developing Concepts 128

3.1.10. Collocational Dimension 10: Social Welfare 132
3.1.11. Collocational Dimension 11: Marking Time and Sectioning 136
3.2. Cluster Analysis of the 11 Collocational Dimensions..... 139
3.2.1. Cluster 1 140
3.2.2. Cluster 2 141
3.2.3. Cluster 3 143
3.2.4. Cluster 4 144
3.2.5. Cluster 5 146
Chapter 4. Discussion and Conclusion..... 148
References 155
Appendix 1 – Tree Tagger Tagset..... 162
Appendix 2 – Python Scripts 163
Appendix 3 – List of the 1,615 nodes and 323 collocates entered into the multivariate exploratory factor analysis 177
Appendix 4 – Numeric expressions used to calculate factor scores 182
Appendix 5 – Descriptive statistics, ANOVA, and means plot from 5-cluster solution HCA for all dimensions 184
Appendix 6 – Dendrograms and Agglomeration Schedule Distance Graphs from HCAs per factor..... 188
Appendix 7 – 5-Cluster solution from HCA for all dimensions..... 199
Appendix 8 – Unique collocations per dimension 203

Lista de Figuras

Figure 1. First and second-order collocates of model: 9a-LogDice(7), R4-L4, C5-NC5; filter: showing collocates with LogDice >8.5…… 30
Figure 2. Example of a text tagged with the Tree Tagger. (corpus file: 1_6_4.txt).. 46
Figure 3. Sample of node-collocate window span for all lemmatized adjectives, nouns, and verbs as potential nodes in the SBE corpus…… 50
Figure 4. Sample of node-collocate window subset for the 2,000 most frequent nodes in the SBE corpus…… 51
Figure 5. Sample of aggregate node and collocate pairs and their frequency in the SBE corpus…… 52
Figure 6. Sample of node and collocate pairs identified in the SBE corpus, showing node and collocate pairs, individual nodes and collocates, their respective POS tags, their joint frequency in the corpus, and the file frequency count…… 53
Figure 7. Strength of association measures on a frequency and exclusivity scale (adapted from Brezina, 2018, p. 74). LL=Log Likelihood; MI=Mutual Information; MI2=Two-way Mutual Information; MI3=Three-way Mutual Information; MS=Minimum Sensitivity; MU=mu-value…… 55
Figure 8. Sample of node and collocate pairs identified in the SBE corpus with their respective logDice values……... 56
Figure 9. Partial dataset of nodes and collocates extracted from the SBE corpus.... 58
Figure 10. Sample of the calculation of mean logDice and number of nodes per collocate for the SBE corpus node and collocate pairs dataset…… 62
Figure 11. Sample of both collocate rankings, showing collocates ranked from highest to lowest according to the count of nodes per collocate…… 62
Figure 12. Results of the Pearson and Spearman's correlations for the two sets of rankings: mean logDice and node count per collocate……....63
Figure 13. Results of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity…….. 64
Figure 14. SPSS v.23 – Factor Analysis: Variable selection…… 65
Figure 15. SPSS v.23 – Factor Analysis: Descriptives…….. 65
Figure 16. SPSS v.23 – Factor Analysis: Extraction…….. 66
Figure 17. SPSS v.23 – Factor Analysis: Rotation…… 67
Figure 18. SPSS v.23 – Factor Analysis – Communalities (sample). (Extraction method: Principal Axis Factoring)…….. 68
Figure 19. SPSS v.23 – Factor Analysis – Total variance explained (sample). (Extraction method: Principal Axis Factoring)…….. 69
Figure 20. SPSS v.23 – Factor Analysis – Scree plot…… 70
Figure 21. SPSS v.23 – Factor Analysis: Extraction – rotated…….. 72
Figure 22. SPSS v.23 – Factor Analysis: Rotation – rotated…… 73
Figure 23. SPSS v.23 – Factor Analysis: Options…… 74
Figure 24. SPSS v.23 – Factor Analysis – Total variance explained (sample) – rotated. (Extraction method: Principal Axis Factoring)…….. 74

Figure 25. SPSS v.23 – Factor Analysis – Pattern Matrix (sample). (Extraction method: Principal Axis Factoring. / Rotation method: Rotation Method: Promax with Kaiser Normalization.)…….. 76
Figure 26. SPSS v.23 – Descriptive Statistics – Standardization……..77
Figure 27. SPSS v.23 – Factor Analysis – Computing factor scores…… 78
Figure 28. Sample of collocations in the SBE corpus comprised of collocates from Factor 1…….. 79
Figure 29. SPSS v.23 – Cluster Analysis – Bivariate correlations…… 81
Figure 30. SPSS v.23 – Cluster Analysis – Linear Regression……. 82
Figure 31. SPSS v.23 – Cluster Analysis – Collinearity diagnostics……. 82
Figure 32. SPSS v.23 – Hierarchical Cluster Analysis…… 83
Figure 33. SPSS v.23 – Hierarchical Cluster Analysis – Statistics and Plots……. 84
Figure 34. SPSS v.23 – Hierarchical Cluster Analysis – Method……. 85
Figure 35. SPSS v.23 – Hierarchical Cluster Analysis – Dendrogram for 1,615 cases (nodes)……... 86
Figure 36. SPSS v.23 – Hierarchical Cluster Analysis – Agglomeration schedule (sample)……. 87
Figure 37. Distance between fusion coefficients in relation to number of clusters…. 87
Figure 38. SPSS v.23 – Hierarchical Cluster Analysis – Save cluster membership.. 88
Figure 39. SPSS v.23 – Comparing means – One-way ANOVA…….. 89

Lista de Tabelas

Table 1. NSF fields and subfields of research study. 43
Table 2. Final composition of the Corpus of Academic Texts (CAT). 45
Table 3. Criteria for collocation identification and extraction. 57
Table 4. Final cluster solutions for each of the HCAs. 92
Table 5. Composition of Dimension 1 95
Table 6. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 1. 96
Table 7. Composition of Dimension 2 101
Table 8. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 2. 101
Table 9. Composition of Dimension 3 105
Table 10. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 3. 106
Table 11. Composition of Dimension 4 109
Table 12. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 4. 110
Table 13. Composition of Dimension 5 113
Table 14. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 5. 114
Table 15. Composition of Dimension 6 117
Table 16. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 6. 118
Table 17. Composition of Dimension 7 121
Table 18. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 7. 122
Table 19. Composition of Dimension 8 125
Table 20. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 8. 126
Table 21. Composition of Dimension 9 129
Table 22. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 9. 129
Table 23. Composition of Dimension 10 132
Table 24. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 10. 133
Table 25. Composition of Dimension 11 136
Table 26. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 11. 136
Table 27. Composition of Cluster 1. 141
Table 28. Composition of Cluster 2. 142
Table 29. Composition of Cluster 3. 144
Table 30. Composition of Cluster 4. 145
Table 31. Composition of Cluster 5. 146

Introduction

The main goal of this study was to propose a bottom-up approach, following Berber Sardinha (2017), to identify and characterize extended collocational networks, or "collocational dimensions", as they are represented in written academic English. It has been established in the literature that collocational networks exist for individual node words hence it would be plausible to say that collocational networks could exist for collocates across nodes, which would be evidenced in a dimension of collocation. Thus, this study set out to extend the concept of collocational networks by identifying the major sets of collocational patterns across nodes, or "collocational dimensions", in academic writing.

The motivation for carrying out this research comes from this researcher's experience teaching English for Specific Purposes (ESP) and English for Academic Purposes (EAP) and working with the English language program of an international network of higher education institutions, where international classes and English-medium classrooms are becoming commonplace. One of the major difficulties of teaching ESP/EAP is identifying the domain-specific or cross-domain lexicon, including collocations, that students may need to be successful in their study or professional area. The focus of most undergraduate teacher training programs is General English and in-service teachers of ESP/EAP need to fend for themselves to find register and domain-specific information to develop their syllabi and/or lesson plans. In general, this means looking for information on the web and using whatever professional documentation their students are able to share. When teaching ESP/EAP, especially when it comes to more specific professional areas, such as for example, Engineering, Health Sciences, Social Sciences, among others, reference books and teaching aids are sparse and not always available in developing countries. Particularly for collocations, the academic collocation lists do not address semantic information nor communicative purposes or situational contexts for the collocations. Recently, when supporting the implementation of a portfolio of international academic solutions, including English language teaching, in higher education institutions in Brazil, it was noticed that several teachers who were experts in the teaching of General English had a hard time inserting ESP into their curriculum to support their students on becoming successful in their future professions. Their main complaint was how

scattered information on domain-specific vocabulary is and how difficult it was to find lexical, multi-word expressions from specific disciplines. This led to the desire of studying collocations and their semantic implications in terms of discipline and register in an effort to identify and characterize the written academic English lexicon. It is hoped that beyond contributing to the body of collocation and lexicon research, the findings of this study may also be used to support the teaching and learning of both ESP/EAP and English medium instruction in undergraduate and graduate classrooms in Brazil and worldwide.

More specifically, the study proposed here aims at contributing to the field of Corpus Linguistics research by applying a novel methodological approach, based on Berber Sardinha (2017), to identify patterns of co-occurrence of sets of collocates across multiple nodes. Therefore, it sets out to extract collocations and analyze the semantic implications of their co-occurrence patterns in written academic English.

Corpus Linguistics may be described as an approach to the study of language and language use, encompassing a range of research purposes and methods, "with an emphasis on language description and theory on systematic observations of language behavior, not on native speaker intuition" (Biber et. al, 2000; Conrad, 2011, p. 49). A corpus can be defined as "a large, principled collection of naturally occurring texts (written or spoken) stored electronically" (Reppen, 2010, p. 2). Thus, Corpus Linguistics studies are based on empirical analysis of authentic texts (i.e., the corpus) with the use of computer programs, incorporating quantitative analysis with qualitative, functional interpretations (Conrad, 2011). According to Hyland (2015, p. 292), "corpus approaches to academic writing provide insights into disciplinary practices which help explain the mechanisms by which knowledge is socially constructed through language".

A central aspect of English—as well as any other language in use—is its inherent variation, that is, the way it is used in different communicative situations (Biber, 1988, 1999). There are innumerable situations in which language is used, just as there are uncountable ways language is configured along a common core of features (Biber, 1999; Hunston, 2002). In terms of academic prose, according to Gray (2015, p. 8), "There is a general consensus, even outside the academic community, that academic writing has distinct characteristics that set it apart from other types of language."

English for Specific Purposes courses, including those for Academic Purposes, have their focus on the skills, language, and genres related to the specific activities needed to be carried out in English. Instructional approaches in English for Specific/Academic Purposes (ESP/EAP) assume that linguistic needs of learners vary according to different language tasks and academic disciplines (Biber, 2006). Vocabulary plays a central role in EAP, and there has been increasing interest in corpus-based research that identifies the major vocabulary items in EAP contexts. Indeed, when defining measures of lexical proficiency, Crossley et al. (2015) ranked the ability to use collocations—"the co-occurrence of two items in a text within a specified environment" (Sinclair, Jones & Daley, 1970/2014, p. 10); the "comings-together-of-words", (Palmer, 1933, p. 1 in Barnbrook, Mason & Krishnamurthy, 2013, p. 24)—as the most important index of vocabulary development.

Although the role of collocation in developing vocabulary competence in EAP has been recognized, previous research on collocations in academic writing has focused on producing collocation lists deemed important for students of English for Academic Purposes to be successful in their studies (Durrant, 2009; Ackermann & Chen, 2013; Lei & Liu, 2018). Those collocation lists have been instrumental in supporting EAP teaching and learning; nevertheless, such lists lack information that is arguably important to language learners such as how the collocations might co-occur with one another and form networks, what general or specific topics they may reflect and be used to talk about, and what registers and/or disciplines they are associated with. Additionally, in Corpus Linguistics studies, collocational patterns have traditionally been studied through concordance lines or KWICs (Key Word in Context), in which lines of text containing the node and its collocates are observed and analyzed (cf. Sinclair et al., 1970/2014), thereby restricting the analysis to individual nodes and their collocates.

The work of Phillips (1985, 1989) and Williams (1998) are notable exceptions. Phillips (1985, 1989) stated that, by analyzing both individual collocations and groups of collocation as well as their recurrent patterns in text, it is possible to identify "lexical networks", described as collocational patterns functioning to lexically structure discourse and give rise to the "aboutness" of a text, that is, its subject matter or textual meaning. Williams (1995) introduced the concept of "collocational networks"— groups of words that share statistical collocates embodying the "multitudinous linkage potential of lexical items" (1998, p. 157). For the author, collocational networks can be

used to identify the lexis of specific discourse communities (what he terms "sublanguage" (1998, p. 156)) since collocations can reveal the meaning of lexical items in texts that are specific to a given discourse community. In addition to the collocation extraction and analyses, Phillips (1985, 1989) and Williams (1998) presented ways of visualizing lexical and collocational networks. Their main reasoning was that "aboutness" or lexical patterns of discourse are revealed through collocational networks which can in turn be visualized through graphs—digraphs in the case of Phillips (1985) and neural network graphs in the case of Williams (1998). A recent computer software (Brezina, McEnery & Wattam, 2015) has expanded on the concepts introduced by Phillips (1985) and Williams (1998) to allow collocational networks to be created and visualized.

While there is much to be gained from being able to extract individual collocational networks and visualize them to analyze collocational patterns in discourse, a major limitation is that it is dependent on the choice of a single initial node to generate the network, which goes back to the traditional analysis of collocational patterns as the examination of the patterns of individual words, often selected ahead of time. That is, the full range of collocations and their networks present in a corpus cannot be identified without choosing, one at a time, a single node word as the starting point of the collocation network. Thus, it became apparent there was a need for a way to verify whether it would be possible to not only extract collocational networks without having to specify a starting node, but also identify collocational networks occurring for multiple nodes at any given time. It is thought that collocates are shared across sets of nodes and they contribute to the lexical structuring of a lexicon, whose patterns could be revealed by extended collocational networks, or "collocational dimensions" (Berber Sardinha, 2017). This study then moves beyond visualization and aims to provide a more comprehensive, empirical, bottom-up view at collocations in academic writing than previously reported, which attempts to quantitatively explore generalizable patterns of co-occurrence of sets of collocates across groupings of nodes.

The corpus used in this study is the Corpus of Academic Texts (CAT), which was compiled for this study. It is comprised of textbooks and journal articles and divided into five subcorpora, one per discipline, according to research discipline areas of the National Science Foundation (NSF): Biological Sciences (BIO); Computer and Information Science and Engineering (CISE); Engineering (ENG); Mathematical and Physical Sciences (MPS); and Social, Behavioral and Economic Sciences (SBE). The

procedures for collection, cleaning, and tagging of corpus texts are detailed in Chapter 2.

Initially, in order to analyze the relationship between collocational patterns and discipline and register, a study was conducted, following a collocational multidimensional analysis approach (Biber, 1988; Berber Sardinha, 2017, Berber Sardinha, Acunzo & São Bento Ferreira, 2016, Berber Sardinha, Ferreira & Mayer, in press), to identify the major sets of interrelated collocations, or dimensions of collocation, in two registers of academic writing across different disciplines. The collocational multidimensional analysis framework was applied to the Social, Behavioral and Economic Sciences (SBE) subset of the CAT and led to the identification of four dimensions of variation across registers and disciplines. The results of the preliminary study can be found in Zuppardi and Berber Sardinha (forthcoming).

Though valuable, the multidimensional analysis framework has the general goal of analyzing register variation. For this study, the desire was to go beyond it to first identify extended collocational networks, or "collocational dimensions", irrespective of register, and, second, to verify whether those collocational dimensions could reflect the structuring of the lexicon in a written academic domain. In order to achieve the stated objectives, the proposed study will attempt to answer the following research questions:

1. When considered across node words in a corpus, is it possible to identify collocational dimensions, that is, can the collocational networks associated with each individual node word be generalized to identify underlying 'dimensions': sets of collocates that tend to co-occur with a set of node words?
2. If so, do those collocational dimensions reflect the aboutness of the lexicon in a written academic discipline domain? To what extent can the collocational patterns in each dimension be interpreted semantically?
3. Do those collocational dimensions also reflect underlying functions in written academic texts?
4. Can scores on the collocational dimensions be used to cluster node words into underlying categories of words that behave in similar ways? If so, how can those clusters of node words be interpreted?

This study attempts to accomplish its goals and answer the established research questions by employing multivariate statistical procedures, such as exploratory factor analysis and cluster analysis. It can be argued that the methodology being proposed provides a way to not only supply a data-driven lexical collocation extraction, but also patterns of collocational co-occurrence across sets of nodes, thereby exploring quantitatively how discourse is woven lexically. In this sense, it is hoped the multivariate analyses will reveal collocational networks that go beyond a single node and its collocates, but rather operate for groupings of collocates across a set of nodes. It is expected that the patterns of co-occurrence among collocates across the most frequent nodes in a corpus will enable the realization of larger collocational associations in the form of collocational dimensions.

This dissertation is organized as follows. The next chapter presents a discussion of the literature relevant to this study and provides a theoretical framework for the methodological approach. The methodology is presented in detail in Chapter 2. Chapter 3 describes the results and their interpretation. Finally, Chapter 4 brings a discussion of the methods and results as well as the main conclusions of this study. The dissertation ends with the references and appendices.

Chapter 1. Literature Review

1.1. Collocations

A widely accepted definition of collocations in corpus-based studies comes from Sinclair, Jones & Daley, who define collocation as

> the co-occurrence of two items in a text within a specified environment. 'Significant' collocation is regular collocation between items, such that they co-occur more often than their respective frequencies and the length of text in which they appear would predict. (1970/2004, p. 10)

In their study, the authors also determine the window span of four words to either side of the word of interest (the 'node') as the "optimum span" to look for collocations. In this case, "[a] 'node' is an item whose total pattern of co-occurrence with other words is under examination; a 'collocate' is any item which appears with the node within a specified environment" (Sinclair et al., 1970/2004, p. 10).

Traditionally, collocations have been studied following the methods established by Sinclair et al. (1970/2014) of looking at collocational patterns through concordance lines or KWICs (Key Word in Context), in which lines of text containing the node and its collocates are observed and analyzed. Sinclair et al. described the "collocational pattern" of a given node as "a list of all words appearing significantly often in its environment [subject to different selection criteria] with information about them" (1970/2004, p. 73), that is, the general patterns of a single node and its collocates, as analyzed through concordance lines. According to Sinclair et al. (1970/2004), meaningful complex relationships between words emerge as a result of significant collocations, providing evidence to what Halliday (1966, p. 153) and Sinclair (1966, p. 426) called "lexical sets"—"group of words with a tendency to occur in the same environment" and express shared concepts or ideas (Sinclair et al., 1970/2004, p. 79); in other words, sets of lexical items determined by their collocational patterns. To illustrate this, Sinclair et al. give the example of lexical items forming significant collocations that express ideas of time:

> The word *time* itself is noticeable for the number of words with a numerical or quantitative connotation which precede it: *first*, *second*, *next*, *last*, *long*, *short*, *all*, *every*, *some*, *full* and *whole*. It is interesting to see that five of these significant collocates are shared with *day*, which also has *one* and *other*, and five with *year*, which also has *third* […]. Evidently words connected with time have a tendency to occur in the same environment. Further examples can be

> seen in the collocational patterns under examination: *ago* is a significant collocate of *two*, *time*, and *years*; *spend* of *time* and *year*; and *after* of *year* and *day*. […]
> Already, from the study of a few nodes, an impression has been gained of a fairly complex network of relationships between the words which are used when expressing ideas of time. No doubt the picture could be extended by an examination of other nodes, and at the same time, the clustering of words related to other areas of meaning would be revealed. (1970/2004, p. 79-80)

As seen in the example above, collocates are shared by different nodes expressing the same or similar ideas, so it is reasonable to say, as predicted by Sinclair et al. (1970/2004), that collocational patterns expressing shared ideas or concepts may occur not only for a single node and its collocates, but also for those collocates across a series of nodes. This supports the need for a different approach to the study of patterns of collocation, not only for individual nodes, but across series of nodes. Although the traditional study of collocations through concordance lines have resulted in significant contributions to different Corpus Linguistics areas, particularly language teaching and learning, those studies have been limited to individual nodes and their collocates.

1.2. Lexical and Collocational Networks

Phillips (1985, 1989) introduced the idea of "lexical networks" as the patterning of collocations that reveal the lexical structuring of texts. Phillips claimed that "it is possible to account for major aspects of the organization of written text by describing its lexical structure" (1989, p. 4). Phillips (1985, 1989) based his claim on the notion of "aboutness" in discourse, that is, large-scale patterning of textual organization may enable awareness in readers of what a text is "about": "[t]he crucial point concerning aboutness is that it is a type of meaning arising from the global structuring of text" (1985, p. 30) and "not a function of particular structures responsible for the local organisation of text" (1989, p. 7). Phillips (1985, 1989) states that the phenomenon of what he defines as aboutness may be objectively observed by looking at individual collocations and groups of collocations (i.e., "sets of lexical items which consistently collocate with each other in text" (1989, p. 5)), and their recurrent patterns in text. According to Phillips (1985, 1989), networks of collocating words reflect the lexical patterns of discourse in a given text which in turn give rise to its subject matter or

“textual meaning” (1985, p. 4), thereby providing evidence that lexical networks are what govern the macrostructure of text.

Halliday and Hasan (1976), discussing text and context in a perspective of language as a social-semiotic phenomenon, express a differing view on what enables awareness of what a text is about. To the authors, the semantics of any language is characterized functionally by four components: experiential, interpersonal, logical, and textual meaning. According to Halliday and Hasan, experiential meaning refers to a text “from the point of view of what it is about—its meaning as the expression of some kind of process, some event, action, state, or other phenomenal aspect of the real world to which it bears some kind of symbolic relation” (1976, p. 18). Thus, experiential meaning expresses the field of discourse, i.e., “the general sense of what [a text] is on about” (1976, p. 24). In this sense, aboutness is seen as a reflection or representation of real-world phenomena and constitutes one component of the semantics of a language. The second component of the semantic system, interpersonal meaning, concerns the participants, mood, “an interaction between speaker and listener” (1976, p. 20)—i.e., the tenor of discourse—, and the third, logical meaning, expresses fundamental logical relations. Textual meaning, the last component, consists of characteristics such as the thematic structure, balance between semantics and grammar, rhythm, and information focus, representing what makes up the texture of a text; in other words, it expresses the mode of discourse, “the particular part that language is playing in the interactive process” (1976, p. 24), for example, if it is written or spoken text. The authors emphasize the view of text as multifunctional:

> The meanings are woven together in a very dense fabric in such a way that, to understand them, we do not look separately at its different parts; rather, we look at the whole thing simultaneously from a number of different angles, each perspective contributing towards the total interpretation. (1976, p. 23)

As aforementioned, Phillips (1985, 1989) approaches aboutness from a different perspective, separating it from the grammatical and logical relations to focus on lexical relationships only. In his view, lexical patterns are what gives rise to perceptions of what a text is about and textual meaning. In order to support his methodology and the claim of lexical networks being what reveals aboutness and the macrostructure of text, Phillips (1985) draws on the conclusions of Sinclair et al., which state:

> This survey has shown that evidence of the lexical organisation of the language can be found by studying the patterns of significant collocation.

> There is thus good reason to expect that intuitively satisfying lexical sets could be formed based on collocational information from a very large body of natural language and using a clustering technique similar to the one developed during the project for identifying the meanings of ambiguous words. (1970/2004, p. 77-78)

Even though the focus of Sinclair et al. (1970/2004) was the phenomenon of collocation per se and the collocational patterns of individual nodes, there is still mention of a possible larger role of lexical sets and their collocational patterns in the structuring of language. Phillips, however, stresses that "[a] lexical set is a paradigmatic concept whilst a lexical network is a syntagmatic one" (1985, p. 100). Thus, for Phillips (1985, 1989), a lexical network implies a notion of sequence, with nodes and collocates occurring on syntagmatic level being the ones to make up a lexical network (possibly because of his claim of lexical networks revealing the macrostructure of text). The present study shares Sinclair et al.'s (1970/2004) view in as much as it considers nodes and collocates that make up patterns of collocations as being part of a lexical or semantic group. It is also important to highlight that although the present study considers the notion of collocational patterns as revealing aboutness of discourse—in addition to semantic fields, topics and subject-matter—, it is not concerned with aboutness characterizing the macrostructure of a text. This study does not concern itself with text structure, but rather lexical patterns of discourse in language.

In order to account for a methodology and theoretical background that could support his argument, Phillips (1985) reviews a large body of conventional linguistic theories concerning lexis, lexicon and semantics (Firth, 1950, 1957; Halliday, 1961; Halliday & Hasan, 1976; Fillenbaum & Rapoport, 1971; Muller, 1977; Katz, 1972; Gruber, 1976; Sinclair, 1983), discourse analysis (Sinclair & Coulthard, 1975; Tadros, 1981), statistical approaches to language studies (Herdan, 1966; Zipf, 1949), collocations (Firth, 1957, 1959; Sinclair et al., 1970), and distributional statistical analysis (Moscovich & Caplan, 1978). He reached the conclusion that none of them by themselves could adequately describe aboutness and macrostructure of text, leading to a choice of what he described as a new quantitative, statistical analysis within a collocational and distributional approach.

The corpus in his study consisted of five university-level science textbooks, two novels, one popular science book, and two randomized books created based on one of the science and one of the non-science books. The ten books were analyzed one

at a time. Phillips (1985, 1989) considered each chapter in a book a “text”, that is, his interval for analysis. Lemmas were used as the unit of analysis, and a collocation was considered as a word pair occurring within a span of four orthographic words of each other to either side with a minimum frequency of at least 2 in the text interval. The methodology consisted of the following steps, carried out for each text interval, i.e., each chapter in the texts that comprised the corpus: (1) text reduction, i.e., eliminating “linguistic carriers of locally organised meaning in text” (1985, p. 53), such as function words (articles, demonstratives, pronouns, pronominal adjectives, modal verbs, auxiliary verbs, conjunctions, interrogative pronouns, interrogative adjectives, prepositions, intensifiers, and interjections) and discourse markers; (2) lemmatization of all remaining items; (3) creation of a wordlist for the text being analyzed; (4) count of the frequencies of collocation for each lemma (i.e., the frequencies of each node and collocate pair) using CLOC (Reed, 1978), a computer program that produces collocations, concordances, and wordlists; (5) relemmatization of the results from CLOC by summing up the frequencies of the individual forms that constitute a lemma—a step needed because the output of the CLOC program is unlemmatized as it reconstitutes the word forms present in the text; (6) configuration of the results in a square matrix format to be used as input for CLUSTAN, a cluster analysis software, where the matrix comprised nodes as rows, collocates as columns, and normalized joint node-collocate frequencies as values; (7) further reduction of the dataset using lower and upper frequency cut-offs ($10 \leq x \geq 300$) followed by random sampling to meet the matrix size restrictions of CLUSTAN of 200 hundred entities; (8) agglomerative hierarchical cluster analysis (Ward’s method) to identify groupings of nodes in the text interval, through the analysis of dendrograms; (9) use of digraph theory (i.e., directed graphs) to map the relations between lemmas and represent lexical networks graphically, where lemmas correspond to the points in the graph and lines represent the frequency of collocation; (10) analysis of the graphs to explore the structure of lexical networks and how they related to the macrostructure of texts; and (11) classification of texts based on the lexical networks identified.

There are several points related to the methodology applied that merit further discussion. First, the text reduction procedure left only lexical items as units to be analyzed with the justification that lexicalized items are the only ones able to reveal aboutness and macrostructure of text. Likewise, the present study adopts the view that only lexical items—rather than grammatical ones—that make up complete, lexical

collocations can reveal topical coherence, shared topics, and semantic fields that make up collocational networks. Another issue concerns the choice for using lemma as the unit of analysis. Phillips (1985, 1989) argues that individual word forms which characterize syntactic organization are not important for the macrostructure of a text, therefore lemmatization becomes necessary. Several studies discuss the choice of using word forms versus lemmas when analyzing collocational patterns, frequently citing the argument that different word forms may have distinct patterns of meaning and reveal varying collocational preferences, which could be disguised by lemmatization (e.g., Hoey, 2005; Sinclair, 1991, 2004; Stubbs, 1996). However, it can be argued, as stated by Evert et al., that "most [lexical] collocations are combinations of lexemes rather than specific word forms" (2017, p. 534); thus, in this study the choice was made to lemmatize all word pairs.

A third, and crucial, point is the use of normalized frequencies in a cluster analysis rather than a measure of statistical association to identify collocations. Phillips (1985, 1989) rejects the use of what he calls the classical model of statistical significance to test the joint occurrence of node and collocate in favor of a non-parametric technique that conceives linguistic data as being spatially organized: "[c]ollocational frequencies can be imagined as constituting a set of data points located at various 'distances' from each other; the more similar the collocational patterning, the closer the relevant points in the analytical space" (Phillips, 1985, p. 71). Hence, he uses node-collocate frequency data to conduct a hierarchical cluster analysis, particularly the Ward's method, with the stated goal of empirically identifying patterns of association among lemmas, namely the syntagmatic lexical networks. Since then, rather than using raw (or even normalized) node-collocate frequency data, it has been an accepted norm to use some type of statistical strength of association to measure the attraction between node and collocate as the input in multivariate statistics procedures, which is the case of the present study, where node and collocate pairs are identified based on a strength of association statistic.

In essence, the cluster analysis conducted by Phillips (1985, 1989) identified groupings of nodes—or clusters—within each of the chapters analyzed. According to the author, each cluster represented a lexical network that could be attributed a semantic label, thereby becoming meaningful constructions of subject matter in text (Phillips, 1985, p. 121-123). Besides identifying the lexical networks through cluster analysis, Phillips (1985, 1989) also represented them graphically as digraphs. This

graphical representation allowed him to further differentiate lemmas in the lexical network. According to Phillips, “[t]he central position of a lemma in a network implies that it tends to attract significant co-occurrences with the largest number of participating lexical items. […] In this sense, the central lemmata can be said to ‘organise’ the networks in which they occur” (1985, p. 141). Phillips (1985) named those central lemmata “nuclear nodes”, noting that most of them appear to be expressions of content, in further support of his argument of lexical networks representing aboutness of text. Phillips (1985, 1989) used chapter summaries, chapter titles, and section headings as a basis for interpreting the lexical networks revealed by the cluster analysis and digraphs. He gives examples of analysis for two chapters per textbook, one of them being chapter four in the ELEN textbook, whose title and section headings are related to operational amplifiers and integrated circuits (Phillips, 1985, p. 126-127). Phillips identifies two major lexical networks centered, unsurprisingly, on the lemmas *amplifier* and *circuit*, and with collocates relating to the content described in the headings. He states that “those […] seem amply to confirm the notion that the lexical network is constructive of topic” (Phillips, 1985, p. 127).

Additionally, the graphs provided information related to connectivity, that is, to what extent the lemmata in each lexical network intercollocate. Phillips (1985) states that while some of the lexical networks did indeed show high connectivity indices, the majority had low connectivity, which he deems evidence that “the tendency of lemmata to collocate with the set of nuclear nodes will far outweigh their tendency to intercollocate” (1985, p. 144). One relevant finding from the analysis of connectivity indices is that those networks that present the highest indices are the ones considered to be highly significant when building subject matter in text, expressing topics that are considered particularly important. Although Phillips’s findings were not substantial, the fact that collocates can and do intercollocate is evidence that construction of meaning in text or subject matter/aboutness may well be supported not only by lexical networks of particular nodes, but also patterns of association among collocates across nodes.

Williams (1998) introduced a way of analyzing collocational patterns through what he termed “collocational networks”, that is, groupings of words sharing statistical collocates that embody the "multitudinous linkage potential of lexical items" (1998, p. 157). Williams (1998) also follows Sinclair’s definition of significant collocation (1970/2004) and defines network as “a web of interlocking conceptual clusters realised in the form of words linked through the process of collocation” (1998, p. 156). His study

proposes collocational networks as a basis for the identification and extraction of lexis pertaining to specific discourse communities, or what he calls the "sublanguage" of a scientific community. Williams (1998) argues that while Phillips's (1985, 1989) notion of aboutness being revealed by collocations could be used to characterize sublanguages, he considers collocation as viewed by Phillips to be "essentially static" (1998, p. 156), where words are pulled together by some force of attraction. Instead, Williams defines his view of collocation as

> an aspect of negotiability in meaning: there is no inevitable collocate, but simply consecration through the frequency of usage. In the expression of a scientific text, the scientist must negotiate meaning through tentative groupings of words that may become stable over time, and hence, collocate or fall by the wayside. What is a 'favourite pair' (Sinclair 1970:122) to the outsider is simply a constructed meaning to the insider. However, rather than rely on a notion of 'pairs' of collocates, it is possible to view collocation as building wider conceptual frameworks, networks of collocates. (1998, p. 156)

Williams (1998) argues that collocations can reveal the meaning of lexical items in texts that are specific to a given discourse community (or discipline), and the conceptual frameworks that guide the language of that discourse community may be visualized through its characteristic collocational networks. Therefore, he posits that lexis changes as the discourse community changes and this can be visualized via collocational networks. Williams states that

> any given word may have a multitude of collocates, and these collocates may have other collocates of their own. This leads to the supposition that if the wordings are seen to reflect the underlying conceptual frameworks, when lexical items are found to co-occur with other lexical items, patterns of co-occurrence will form "collocational chains". Following these "collocational chains" through would allow us to go beyond the immediate "contextual framework" to isolate the full frame of reference of a given item within the lexis of a given discourse community. (1998, p. 156)

The author illustrates the concept of collocational networks using a corpus of plant biology research articles divided into 3 subcorpora, namely: (1) molecular biology; (2) physiology of parasitic plants; and (3) interdisciplinary research in parasitic plants. He looked at collocations that are comprised solely of lexical items (orthographic word forms) and whose strength of association is determined by the Mutual Information (MI) score. A collocational chain starts by taking a single initial node and its collocates, determined by MI scores with cut-off points at 4, 5, and 6. Each of those collocates then becomes a new node and the network of its collocates is added. The collocational chain continues to grow as all nodes and collocates are gradually added until such a time as no more significant collocates are found. Williams

(1998) selected the first 50 lexical word forms in the frequency list of each subcorpus as the initial nodes for analysis in the collocational chain. Collocates were extracted for each of those nodes and MI was then calculated "for collocates with a frequency of co-occurrence of 8" (1998, p. 157). Following this approach, Williams identifies what he calls "nuclear nodes", that is, those nodes that "may have a higher concentration of collocates, representing concepts central to the subject field" and whose centrality is indicative of them being surrounded by clusters of other nodes (1998, p. 158). From his description, the distinction between "collocational chains" and "collocational networks" is unclear. It is assumed that "collocational networks" are the visual representations of what Williams (1998) operationalizes as "collocational chains". Additionally, a number of decisions made during the extraction of nodes and collocates are unclear, such as: the exact window span, or collocation window; the choice for MI over other strength of association measures; the restriction of collocate co-occurrence frequency of 8; and how nuclear nodes were identified. Unlike the digraphs used by Phillips (1985, 1989), Williams (1998) proposes the concept of neural network mapping to illustrate collocational networks graphically (Kohenen Self-Organizing Maps–Faussett, 1994, p. 429 in Williams, 1998, p. 158,168-169). The article showcases the collocational networks for the nuclear nodes *DNA* and *plants*. The neural network graphs illustrate the collocates of those nuclear nodes as well as their relationships to other nuclear nodes. Williams argues that this is evidence of the conceptual frameworks of a given sublanguage (or discourse community) being "lexically realised through closed set collocational networks" (1998, p. 160).

1.3. Visualizing Collocational Networks

Several other studies have explored the idea of texts in different fields of discourse or disciplines being "organised into lexical patterns, which can be visualised as networks of words that collocate with each other" (cf. Brezina, McEnery & Wattam 2015, p. 142-143). While studying discourses related to swearing in English, McEnery (2006) makes use of collocation networks as part of his analysis. The starting nodes were chosen based on a keyword analysis and the collocates were identified based on MI2 (Mutual Information squared) score above 3 and a window span of ± 5.

McEnery (2006) also considered directionality when creating the collocation networks, that is, whether the collocate was to the right or to the left of the node.

Recently, Brezina, McEnery and Wattam (2015) developed a specific computer software with the aim of graphically illustrating collocation networks, expanding on the concepts identified by Phillips (1985, 1989) and Williams (1998) and, in particular, in an effort to replicate and add to the findings of the McEnery (2006) study. In a sense, although following different methodologies, the collocation networks being created by their software appear to be fulfilling the same concepts as "collocational chains" associated with an individual node from Williams (1998). Their software builds collocation networks based on user-defined corpora and a multitude of statistical association measures and shows graphically—through symbols and connecting lines—the relationships among nodes and collocates. Brezina et al. (2015) emphasize the importance of taking different criteria into account when building collocation networks, such as: (1) collocation window, or window span, between node and its collocates; (2) node and collocate frequency of co-occurrence; (3) exclusivity, as measured by strength of association statistics; (4) directionality, using Delta P as a measure; (5) dispersion, i.e., "the distribution of the node and the collocates in the corpus" (2015, p. 140); (6) type-token distribution among collocates, as measured by the lexical gravity *G*; and (7) connectivity between individual collocates. The first three criteria are well-established in the literature for collocation identification. Criteria 4-6 are proposed by Gries (2013), and the last item was added by Brezina et al. (2015). The concept of connectivity as defined by the authors is of relevance to this study in as it states that "[c]ollocates of words do not occur in isolation, but are part of a complex network of semantic relationships which ultimately reveals their meaning and the semantic structure of a text or corpus" (Brezina et al. 2015, p. 141). The study being carried out here goes further to argue that both nodes and collocates cannot be considered in isolation but must be taken as part of a larger set of collocational networks, namely collocational dimensions. Thus, we can assume that collocates inter-collocate and this inter-collocation leads to larger, extended collocation networks.

Both Williams (1998) and Brezina et al. (2015) claim their studies to have a basis on Phillips's notion of "lexical networks" (1985, 1989), who, as aforementioned, attempted to describe "aboutness" of discourse in terms of collocational patterns and how it relates to the macrostructure of text. However, different methodologies notwithstanding, unlike Phillips (1985, 1989), Williams (1998) and Brezina et al. (2015)

are not concerned with text or “aboutness” of discourse. Looking at collocational networks based on a single initial node does not provide information on the “aboutness” of discourse or how that discourse is lexically structured in text. What the networks do provide is a description of collocational patterns for individual nodes and their collocates. In a nutshell, Phillips (1985, 1989) was concerned with lexical patterns in texts and Williams (1998) and Brezina et al. (2015) were concerned with lexical patterns associated with node words regardless of text boundaries. It is possible to say that Williams (1998) and Brezina et al. (2015)’s work are an extension of what Sinclair et al. (1970/2004, p. 73) named a “collocational pattern” in as much as they are concerned with looking at the patterns of nodes and their first-, second-, third-order collocates, and so on in order to achieve collocational chains/collocation networks. While the graphical representation of collocation networks is of great significance in describing collocational patterns in discourse, it is still dependent on the researcher or software end-user choosing a single initial node to be analyzed and on the rendering capabilities of the software. In other words, the software does not identify the range of collocations and their networks present in the uploaded corpus, rather it shows, one at a time, the collocation network for an arbitrary single node word as a starting point, which must be inserted by the user as a search word. In all aforementioned studies, collocational networks have been built based on a single node word, which may be identified through varying methods and statistics but is still indispensable to create the initial collocational network. This in effect renders non-generalizable graphs, that is, the patterns illustrated by the collocation networks of a given node cannot be generalized across different sets of nodes and collocates because it always requires a new node word to be selected—be it an initial node or a collocate from the network—in order to build the next network for that word. Take as an example the following collocation network for the noun *model*, the most frequent lexical item in a corpus of 200 journal articles of Social, Behavioral and Economic Sciences (Figure 1), built using the software developed by Brezina et al. (2015). It shows the collocation network for the initial node *model*, i.e., the collocates with high strength of association. In order to see second-order collocates, you must choose one of those collocates to discover the next collocation network. In this case, the collocate *data* was chosen as the next node. The graph does show which collocates are shared between the two nodes, but it is limited to that. To see a larger pattern, a new collocate must be chosen as the next node.

Figure 1. First and second-order collocates of model: 9a-LogDice(7), R4-L4, C5-NC5; filter: showing collocates with LogDice >8.5.

Taking a different approach, but still with a focus on visualization, Mollet, Wray and Fitzpatrick (2011) used lexical co-occurrence networks to visualize what they termed ‘second-order collocations’, “notable patterns of behaviour by one word within the context of another word’s co-occurrence patterns” (2011, p. 87). Their study attempted to go beyond a single initial node and its collocates, trying to map expanded collocational networks. The authors employed a network modelling algorithm to extract lexical items and calculate the weight of the connection between them so that networks could be visually represented. In their study, network is defined as “consist[ing] of a collection of nodes connected by lines” (2011, p. 91), where a collocation “refer[s] to the relationships that display […] an association that seems to imply a semantic association” (2011, p. 115). Their algorithm uses a window span of ± 4 content words to identify lexical items that are linked. This means that the actual window span may be larger than the traditional ± 4 orthographic words, since it continues to search the window until it has identified 4 content words to either side of the node. Rather than using a statistic of association to measure the strength of association between the node and its potential collocates, the authors measured the weight of the connection in terms of “the reciprocal distance in content words 1/distance” (2011, p. 93), where distance is the number of words between the node and the content word. Then, the Frequency Normalized Weighted Distance (FNWD) was calculated for each node and its collocates to indicate their patterns of co-occurrence. “[T]he extent to which words

that are each connected to the focal item are also connected to each other" (Mollet et al., 2011, p. 94) is given by the curvature of the node in the network. According to the authors, high curvature values indicate that the lexical item tends to co-occur with the same words. The network is then represented graphically.

Mollet et al. (2011) exemplify the method using two lines of text from Jane Austen and then presents a case study of how the word *order* as a second-order collocate may affect the meaning of the lexical item *social* and the tag *<formulae>* in the British Academic Written English Corpus (BAWE). The authors built the individual graphical networks for *social* and *<formulae>*, then selected *order* as the next node, thereby showing only the collocates that are shared between *order* and the lexical item (either *social* or *<formulae>*). Although different methods are used to extract word pairs, Mollet et al.'s (2011) representation of second-order collocates is the same as the one achieved by the software developed by Brezina et al. (2015), as illustrated above. The authors acknowledge the possible limitations of the computationally-demanding methodology and provide suggestions of applications to word meaning, stylistics, critical discourse analysis, and linguistic competence studies. According to Mollet et al., "[w]hat second-order collocation measurements reveal is something about a word's location in cognitive space – space that is determined not only by the behaviour of the focus word, nor only of its collocates, but also of its collocates' collocates" (2011, p. 113). The study contributes to the understanding that collocational patterns go beyond individual nodes and their collocates and highlights the need for patterns of co-occurrence of collocates to also be taken into account as constructive of meaning in discourse. However, the limitation of having to specify an initial node or target item remains.

1.4. Dimensional Collocational Analysis

The relationship between the node and its collocates in collocation network graphs are clearly shown in the previously mentioned studies (Williams, 1998, Brezina et al., 2015, Mollet et al., 2011), but what is missing is the relationship between the collocates themselves across a series of nodes. It is possible to see somewhat that kind of relationship when you chose one or more collocates as the next node in the network

(as shown in Figure 1), but that is still limited to the specific node chosen. That is, it is not possible to see a generalizable pattern of how collocates are co-occurring across nodes in the corpus. You must start by choosing a node, exploring the collocates for that node, then picking a collocate that will in turn become a node whose collocate pattern will be analyzed, and so on. Phillips (1985) referenced this when discussing the issue of choosing a node as an initial point of analysis: “the collocational patterning of all nodes must be examined simultaneously in order to determine on as objective a basis as possible which nodes are to be considered as initiating patterns of association or, indeed, whether the notion of initiation is in fact appropriate” (1985, p. 66). The goal of building collocational networks across different sets of nodes and collocates should be to look at the way discourse is lexically structured, but a collocational network graph does not allow for that question to be answered since it looks at a single word at a time, no matter how principled the choice of an initial node may be.

This study aims to fill this gap, by proposing a bottom-up approach to identifying and characterizing extended collocational networks, or “collocational dimensions”, as they are represented in academic discourse. It is expected that collocates are shared across sets of nodes and they contribute to the organization of a lexicon, whose patterns could be revealed by the collocational dimensions. It has been established that collocational networks exist for individual node words hence it would be plausible to say that collocational networks could exist for collocates across nodes, which would be evidenced in a dimension of collocation. Thus, this study sets out to extend the concept of collocational networks by identifying the major sets of collocational patterns across nodes, or “collocational dimensions”, in academic writing. This moves beyond visualization, which is not the goal of this study. Rather, it is a more comprehensive, empirical, bottom-up view at collocations in academic writing than previously reported, which attempts to quantitatively explore generalizable patterns of co-occurrence of sets of collocates across groupings of nodes.

1.4.1. Theoretical Considerations

1.4.1.1. Lexicon Organization and Semantic Analysis

One of the goals of this study is to analyze whether collocational dimensions reflect the structuring of lexicon and aboutness/meaning relations within them.

Organization of a lexicon, specifically in terms of sense relations, is a widely studied concept in natural language processing (NLP), where a lexicon is often seen as "a list of words in a language—a **vocabulary**—along with some knowledge of how each word is used. A lexicon may be general or domain-specific" (Hirst, 2009, p. 269). It usually contains content words (such as nouns, verbs and adjectives) as well as multi-word expressions, such as collocations, phrasal verbs or fixed phrases. A lexical entry, each word or phrase in the lexicon, specifies a number of properties attributed to the word or phrase, including spelling or sound, grammatical behavior, aspects of use, meaning, and relationships with other words (Hirst, 2009).

The most widely known example of a computational lexicon is WordNet (Fellbaum, 1998a), a semantic database that was designed to model the human mental lexicon of English and whose basic lexical entry is the synset (or synonym set), comprised of groupings of words with closely related meanings (i.e., word senses). According to Fellbaum (1998b), in WordNet, "the meanings of nouns, verbs, adjectives, and adverbs are represented in terms of their links to other (groups of) words via conceptual-semantic and lexical relations. Each part of speech is treated differently reflecting different semantic properties" (1998b, p. 209). While discussing the relationship between lexicons and ontologies, Hirst (2009) states that

> [b]y definition, a **word sense**, or the "meaning" of a word, is a semantic object—a **concept** or **conceptual structure** of some kind, though exactly what kind is a matter of considerable debate, with a large literature on the topic. Among other possibilities, a word sense may be regarded as a purely **mental object**; or as a structure of some kind of **primitive units of meaning**; or as the **set of all the things in the world that the sense may denote**; or as a **prototype** that other objects resemble to a greater or lesser degree; or as an **intension** or **description** or **identification procedure**—possibly in terms of necessary and sufficient conditions—of all the things that the sense may denote. (2009, p. 273-4)

Regardless of how one defines word sense, the idea that holds is that meaning(s) may be identified and classified through semantic analysis, both for individual lexical items as well as for relationships among lexical items. In this study, it is expected that collocational dimensions may reflect word senses or meaning relationships among groups of co-occurring nodes and collocates.

The lexicon, its structure and semantic relationships, is the basis for NLP applications concerning sense relations and semantic analysis. According to Poesio (2000),

> The ultimate goal, for humans as well as natural language-processing (NLP) systems, is to understand the utterance […]. 'Understanding' an utterance is a complex process, that depends on the results of parsing, as well as on lexical information, context, and commonsense reasoning […]. (2000, p. 93)

This goal may be equaled, in NLP interpretation, to semantic analysis of phrases, sentences, individual words, fixed expressions (such as collocations), and utterances in context (Goddard & Schalley, 2010). Goddard and Schalley (2010) mention a few of the NLP applications drawing on WordNet, e.g., information retrieval and extraction, disambiguation, categorization of texts, as examples of the demand for studies on lexical relations and linguistically informed ontologies.

Distributional semantic modeling, a type of semantic analysis in NLP, "contends that it is possible to approximate what humans do when they learn word meanings via similarity judgements […] by operationalizing semantic similarities between linguistic units based on their distributional properties" in large corpora (Desagulier, 2019, p. 1). In other words, it may be possible to infer the meaning of a word from its distribution in multi-vector space. The idea of meaning being derived from usage is not new and the theoretical basis for distributional semantics may be found in Firth's famous quote "you shall know a word by the company it keeps" (Firth, 1957, p. 11). Firth (1957; 1968) described meaning as being dependent on the utterance, the participants, and the context of situation, that is, "[w]ords must not be treated as if they had isolate meaning and occurred and could be used in free distribution" (Firth, 1968, p. 18). Instead, meaning is built from the words in its surroundings and the context in which it is used.

Distributional semantic models are accomplished by transforming the distributional properties into vectors (i.e., strings of real numbers), where "words with similar vector representations have similar meanings" (Desagulier, 2019, p. 1). The most notable current example of this is word2vec (Mikolov et al., 2013), a group of related distributional models used to generate word embeddings. It can be used with either a continuous bag-of-words (CBOW) or continuous skip-gram architecture. Although word2vec is currently considered to be the algorithm with the most advantages for producing state-of-the-art word embeddings, the reasons why it works so well are not yet completely understood. (Levy et al., 2015, p. 211).

The study reported herein does not have as its primary goal a full-scale semantic analysis as those carried out in NLP studies. However, the it does have the similar goal of analyzing sense relations and meaning (i.e., aboutness) reflected by the collocational patterns identified in the collocational dimensions and possibly discovering to what extent they reflect the discipline domain and written academic discourse, albeit following a different methodology than that of NLP studies. Additionally, this study aims to contribute to studies concerning the organization of the lexicon by revealing how the lexicon is organized with respect to collocations.

While semantic analysis and distributional semantics in NLP have the goal of grouping words that behave similarly by looking at the semantic relationships between those words, the methodology being proposed here has the goal of grouping collocates, a special function of words, with respect to their co-occurrence across another set of groups of words, the nodes. The grouping of words according to their sense relations is accomplished in this study as a secondary step, by grouping the nodes and looking at the semantic relationships between them with respect to the collocational dimensions.

1.4.1.2. Language Functions

Traditionally, collocation research has analyzed and described how a node tends to co-occur with a set of words, i.e., its collocates, due to the nature of that node. The idea proposed here is that beyond the traditional notion of a collocation exists a tendency for preference among the collocates themselves because they are inter-related, apart from the nature of an individual node, such that they share underlying functions in discourse. The role of function in discourse has been discussed in the literature. Halliday and Hasan (1976), going beyond the traditional interpretations that argue functional variation equals variation in language use (Malinowski, 1923; Buhler, 1934; Britton, 1970; Morris, 1967 apud Halliday & Hasan, 1976), describe function “not just as the use of language but as a fundamental property of language itself, something that is basic to the evolution of the semantic system” (1976, p. 17). As aforementioned, Halliday and Hasan (1976) argue that functional interpretations of the semantics of a language depend on four strands of meaning—experiential, interpersonal, logical, and textual—which are “interwoven in the fabric of discourse” (1976, p. 23). Biber (1995),

while discussing the theoretical issues related to the multi-dimensional analysis framework—a methodology which inspired the methods proposed in this study, as seen in the next subsection—, states that “linguistic variation is typically conditioned by some combination of social, situational, discourse, and processing ‘functions’” (1995, p. 136). The author describes four major types of function of linguistic forms, that can be present independently or combined: (1) tasks performed in discourse by the linguistic form; (2) situational characteristics reflected by the linguistic form; (3) constraints in production circumstances reflected by the linguistic form; and (4) situational or social group distinctions conventionally indexed by the linguistic form (Biber, 1995, p. 136). In this study, it is expected that collocational patterns across nodes identified in the collocational dimensions may possibly reflect shared underlying functions in discourses, where collocations may serve particular discourse tasks, reflect situational/production characteristics and/or communicative purposes, and/or reflect conventional markers of the texts in the corpus.

1.4.2. Methodological Considerations

The methodology being proposed in this study is inspired by the collocational variant (Berber Sardinha, 2017; Berber Sardinha et al., forthcoming) of the multi-dimensional framework introduced by Biber (1988). In MD studies, the goal is to identify the underlying parameters of variation in the data. In a traditional functional multi-dimensional analysis, the goal is to identify patterns of variation in texts across different registers based on co-occurrence of lexicogrammatical features. Texts are the units of observation and linguistic features present in the corpus are the variables. The texts being analyzed are tagged for lexicogrammatical features and normed counts of the features are submitted to a statistical procedure, namely factor analysis. In a factor analysis, the variables—in this case, the linguistic features present in the corpus—are reduced to a limited number of factors comprised of the set of variables that co-occur in the texts. Each factor is then analyzed qualitatively to identify a functional dimension that reflects one or more shared communicative functions underlying the patterns of co-occurrence of the linguistic features that comprise that factor. The final result is a set of dimensions that reflect the underlying patterns of variation in linguistic features across the registers being analyzed.

Similarly, the goal of a collocational MD analysis is also to identify patterns of variation; however, rather than analyzing the co-occurrence of linguistic features, it aims to determine the patterns of co-occurrence and the underlying use of collocations across text varieties. Berber Sardinha (2017) developed this variant of the MD analysis in order to verify whether there are register-specific collocations and if register awareness could be predicted by those collocations. His argument was based on Hoey's (2005, 2013) theory of lexical priming, which claims that information about registers—or textual varieties—in which specific collocations occur are stored in the speaker's mind. According to Berber Sardinha, "if presented with a particular collocation, individuals should be able to identify the register for which it is characteristically primed. This hypothesis actually presupposes a regular association between collocation and register, meaning that register differences should be marked by differences in collocational use" (2017, p. 189-190). Thus, Berber Sardinha (2017) conducted a collocational multidimensional analysis on the full-text, downloadable version of the Corpus of Contemporary American English (COCA, 1990-2012). He also followed Sinclair et al.'s definition of significant collocation (1970/2004), identifying and extracting word pairs within a window span of ± 4 words. The methodology used can be summarized as follows: (1) identification and extraction of the 300,000 most frequent word pairs in each of the five registers present in COCA; (2) from those, identification and extraction of the 2,000 most frequent lemmatized nouns, main verbs, and adjectives and their collocates in each register; (3) computation of the strength of association between each node and its collocates using the logDice statistic; (4) merging of the lists of word pairs per register into a single spreadsheet with nodes as variables, collocates as the observations, and the logDice statistic as cell values; (5) initial and rotated factor analysis, with Principal Axis Factoring as the extraction method and Promax rotation, with nine factors being extracted; (6) calculation of factor scores for each collocate in the data set; (7) interpretation of the nine-factor solution; (8) naming of the dimensions identified; (9) calculation of mean dimension scores in order to graph results; and (10) statistical tests (F-score and R^2) to measure the extent of variation across registers. Steps 1-4 and 7 were accomplished with computer scripts written specifically for the study.

The interpretation of the factor solution results is what differs the most from a traditional, functional MD analysis. In a functional MD analysis, interpretation is based on analysis of the functional and communicative purposes fulfilled by the

lexicogrammatical features loading in each factor. The interpretation carried out by Berber Sardinha (2017) in his collocational MD analysis involved, in addition to a series of other techniques, analyzing the collocations in each factor and tagging text excerpts with the USAS (UCREL Semantic Analysis System) English semantic tagger to identify the most prevalent semantic fields to which the words belonged. According to Berber Sardinha,

> the interpretation of the factors as dimensions took into consideration a range of constructs, tools, and techniques, including the whole set of collocations in the factor, the semantic preferences exhibited by the collocations, their lexical or semantic fields, the 'aboutness' manifested in the items, the subject matter and topics expressed, the major collocations sorted by their factor scores, KWICs of node–collocate combinations (obtained from the online version of the corpus), the variation (or lack thereof) across the registers, and samples of texts in which the collocations occurred. (2017, p. 196)

The author identified nine dimensions of collocation in American English, namely: (1) Literate discourse; (2) Oral discourse; (3) Objects, people, and actions; (4) Colloquial and informal language use; (5) Organizations and the government; (6) Politics and current affairs; (7) Feelings and emotions; (8) Cooking; and (9) Education research. Finally, Berber Sardinha (2017) conducted a discriminant function analysis to verify the extent to whether register categories can predict collocational use. The study showed that collocations do reside on a multidimensional space, just as texts, where each dimension represent "a distinct semantic parameter underlying the use of collocations in text" (Berber Sardinha, 2017, p. 224). In addition, the study found that collocation was generally a poor predictor of register differences in general English, registers in turn were good predictors of collocation (Berber Sardinha, 2017).

Berber Sardinha et al. (forthcoming) also employed a collocational MD analysis with the aim of creating a dictionary of collocations based on the Brazilian Corpus, a one-billion-word corpus of spoken and written Brazilian Portuguese. The authors followed a similar methodology as Berber Sardinha (2017), with the main difference being the reversal of variables and observations, that is, for the dictionary, nodes were the observations and collocates were variables. Berber Sardinha et al. (forthcoming) used a Python script on the Sketch Engine API to: (1) identify the 10,000 most frequent words in the corpus, which were considered as nodes after cleaning; (2) run up to 20,000 random concordance lines for each of the nodes; (3) extract a collocate table for each concordance line, sorted by logDice; (4) save the top 50 collocates from each table as a set; (5) identify the 2,000 most frequent collocates across the sets; and (6)

merge the sets in a single dataset. The dataset was then submitted to a factor analysis, resulting in the identification of eight dimensions of collocation. Most of the dimensions of collocation identified reflected different discourse domains; however, a few captured functional communicative purposes being fulfilled by the collocations, namely dimensions 1 (Hedging and cognition), 3 (Expressing likelihood), and 4 (Oral language versus specialized internet usage and conventions). The remaining dimensions were: Dimension 2, Crops, livestock, and nutrition; Dimension 5, Science and technology; Dimension 6, The economy versus emotions; Dimension 7, Medical discourse versus politics, civil service, and government; and Dimension 8, Crime and law enforcement.

Although inspired by the collocational MD analysis developed by Berber Sardinha (2017) and Berber Sardinha et al. (forthcoming), as aforementioned, a major difference between the methodology in the present study and previous MD studies—functional, lexical, and collocational—is the lack of comparison across texts and registers. Though it could be an intended extension and future direction, this study attains itself to detecting patterns of variation of node and collocate pairs in the whole corpus, irrespective, at this time, of individual texts or specific registers. The methodology proposed here makes use of: (1) dimension reduction procedures (i.e., multivariate statistics) to reveal underlying patterns of variation of collocates co-occurring across nodes; and (2) cluster analysis to classify nodes into categories in relation to the shared collocational patterns identified in the dimensional collocational analysis, with the aim of helping interpret the dimensions and finding evidence of extended collocational networks for groups of nodes. Therefore, in this study, nodes are the units of observation, rather than texts and/or registers, with collocates as the variables entered into the factor analysis. This means that both variables and units of observation are the same items—i.e., they are linguistically the same: nouns, verbs, and adjectives—since nodes can be collocates and vice-versa. The extent to which this may impact the results is yet unknown. Factors then are groups of co-occurring collocates across nodes and collocational dimensions represent the underlying variation of collocational patterns—realized as extended collocational networks—found in the corpus.

In a way, each row of observation in the data being submitted to the factor analysis can be considered the collocation network for the node being observed in that row, since the row comprises the collocates of each particular node. That is, the row shows the node of interest and the strength of association of all collocates for that

node. Therefore, each row shows a separate collocation network. By performing a factor analysis, in addition to the relationships between individual nodes and their collocates, the relationships between collocates themselves are also revealed. It could be argued that, in the visualization studies discussed above, the second-, third-order collocation networks and so forth in a way reveal those relationships, but they are not generalizable across nodes and collocates since they must always start from a chosen node or collocate-turned-node. Moreover, as mentioned above, the goal of this study is not visualization, but instead, the empirical identification of co-occurrence patterns among collocates across different sets of nodes. It is hoped that a collocational dimension will be able to measure empirically how a set of collocates co-occur across different nodes, showing a collection of collocates, or an extended collocational pattern, that exists across those nodes.

The interpretation of the factors is carried out similarly to the interpretation done in an MD analysis. The interpretation in this study will be mostly based on: (1) the co-occurring collocates in a given factor; (2) analysis of collocates with respect to particular node words; and (3) the distribution of the sets of collocates across groupings of nodes. In addition, the collocations found in each factor are also analyzed for their semantic preferences, lexical or semantic fields, as well as aboutness, topics and subject matter expressed. As aforementioned, lexical items and collocations may reveal the aboutness and fields of discourse of a text (Halliday & Hasan, 1976; Phillips, 1985, 1989; Williams, 1998). In this study, it is argued that aboutness, topics, and subject matter can also be constructed based on the extended collocational patterns that structure the lexicon, that is, the collocational dimensions.

In order to support the analysis of the distribution of the sets of collocates across groupings of nodes as well as the characterization of the role those extended collocational patterns play in the organization of the lexicon and in weaving discourse lexically, it is essential to analyze how the nodes are grouped in terms of the dimensions. In other words, nodes may be grouped according to similar collocational preferences as captured by the dimensions. To that end, a cluster analysis may be employed to drive a bottom-up identification of “node types” or categories in relation to the collocational patterns identified by the collocational dimensions, thereby furthering the interpretation of extended collocational patterns in academic discourse. The factor scores of collocates in each collocational dimension are then the quantitative variables that allow nodes that behave in similar ways to be grouped into

categories. According to Staples & Biber, “the goal [of a cluster analysis] is to create a new categorical variable that minimizes the amount of variation within categories, while maximizing the differences among categories” (2105, p. 244). In this study, it is expected that node types will emerge as new categorical variables resultant from the cluster analysis.

It is hoped that the methodological approach proposed here will provide evidence of the existence of extended collocational patterns and how they are lexically woven in discourse, thereby contributing to the fields of Corpus Linguistics and theories of meaning, as well as to the body of lexical and collocational research.

The next chapter presents the methodological decisions and steps carried out in this study.

Chapter 2. Methodology

This chapter describes the methodological approach used in this study. Overall, the methodology undertaken comprises the following steps: (1) use of principled and adequate corpora for the research being carried out; (2) use of computer programs to: convert and clean texts, add part-of-speech tags to all texts, extract node and collocate pairs, count their frequency in the corpus, and calculate the strength of association between node words and their collocates; (3) use of statistical techniques to identify collocational dimensions and node types; and (4) use of qualitative analysis of the underlying parameters of variation that account for the extended collocational networks identified through quantitative methods. Particularly, the methodology applied in this study follows these steps:

A. Corpus design, compilation, and tagging
B. Collocation identification and extraction
C. Exploratory factor analysis for dimensional collocational analysis
D. Cluster analysis
E. Qualitative analysis and interpretation

Collocation identification and extraction as well as factor and cluster analyses were carried out in order to identify the collocational dimensions and node types of written academic English, as detailed in the methodology sections below and in the results described in Chapter 3.

2.1. Corpus Design and Compilation

The Corpus of Academic Texts (CAT) was collected specifically for this study to reflect two written academic registers: textbooks and journal articles. It is a monolingual corpus, divided into 5 subcorpora, one per discipline area, according to the National Science Foundation (NSF)[1]: Biological Sciences (BIO); Computer and Information Science and Engineering (CISE); Engineering (ENG); Mathematical and Physical Sciences (MPS); and Social, Behavioral and Economic Sciences (SBE). For each

[1] National Science Foundation (NSF) research areas can be found at https://www.nsf.gov/about/research_areas.jsp.

discipline, 30 textbooks and 200 research articles were collected, totaling 50,865,265 tokens (3,165,286 types) across 1,150 texts. In this study, register is understood as "a cover term for any language variety defined by its situational characteristics, including the speaker's purpose, the relationship between speaker and hearer, and the production circumstances" (Biber, 2009, p. 823); a discipline domain is understood as a research area, including its fields and subfields, as defined by the NSF.

The following NSF fields and subfields of research study were used as guidance in the collection of written texts (Table 1). The NSF classification was used because it follows a widely recognized system for classifying research areas in a number of disciplines.

Table 1. NSF fields and subfields of research study.

Engineering (ENG)	**Mathematical and Physical Sciences (MPS)**
Chemical, Bioengineering, Environmental and Transport Systems (CBET) Civil, Mechanical and Manufacturing Innovation (CMMI) Electrical, Communications and Cyber Systems (ECCS) Emerging Frontiers and Multidisciplinary Activities (EFMA)	Astronomical Sciences (AST) Chemistry (CHE) Materials Research (DMR) Mathematical Sciences (DMS) Physics (PHY)
Biological Sciences (BIO)	**Social, Behavioral and Economic Sciences (SBE)**
Environmental Biology (DEB) Integrative Organismal Systems (IOS) Molecular and Cellular Biosciences (MCB)	Behavioral and Cognitive Sciences (BCS) Social and Economic Sciences (SES)
Computer and Information Science and Engineering (CISE)	
Advanced Cyberinfrastructure (ACI) Computing and Communication Foundations (CCF) Computer and Network Systems (CNS) Information and Intelligent Systems (IIS)	

Textbooks were collected based on the Amazon Best Sellers lists for each of the disciplines. The Amazon Best Sellers lists from December 2016 – July 2017 were used. The following textbook categories and subcategories were included in the search for candidate textbooks for inclusion in the corpus:

A. Business & Finance: Economics

B. Computer Science: Artificial Intelligence; Database Storage & Design; Networking; Object-Oriented Software Design; Operating Systems; Software Design & Engineering
C. Engineering: Chemical Engineering; Civil Engineering; Electrical & Electronic Engineering; Environmental Engineering; Industrial Engineering; Mechanical Engineering
D. Medicine & Health Sciences: Medicine - Biotechnology
E. Science & Mathematics: Astronomy & Astrophysics; Biology & Life Sciences; Chemistry; Earth Sciences; Environmental Studies; Mathematics; Mechanics; Physics
F. Social Sciences: Anthropology; Political Science; Psychology; Sociology

In total, 120 textbooks–30 per discipline–were collected in pdf or epub formats and converted to text files. For every textbook, only introduction and chapter content were kept, including: tables, figures, key terms, exercises, cases studies, and endnotes. Text files were marked with the following tags: (1) /TXTBEGIN/, preceding introduction or chapter 1 content, and (2) /TXTEND/, following the last sentence of the last chapter in the textbook. A computer script was used to extract the text within those tags.

Journal articles were collected from top ranking journals in each discipline as classified by the Scimago Journal & Country Rank (SJR).[2] The SJR indicates a journal's influence as measured by number of citations and prestige of journals from where those citations come. Journals were chosen from the Journal Rankings from June 2017 – December 2017. Selection criteria included: being in the top 100 journals; having subject area and category match with an NSF discipline area and/or subarea; and having articles published in English only. In total, 800 articles–200 per discipline–were collected in pdf format and converted to text files. For every article, boilerplates, acknowledgements, references, and supplemental materials were excluded. The following were kept: abstract, keywords, full article content, end-of-article methods section (when applicable), footnotes and/or end-of-article notes (when present), and glossaries (when present). Text files were marked with the following tags: (1) /TXTBEGIN/, preceding abstract or introduction, and (2) /TXTEND/, following the last

[2] The Journal Rankings by the Scimago Journal & Country Rank (SJR) can be found at https://www.scimagojr.com/journalrank.php.

sentence of the last section in the article. A computer script was used to extract the text within those tags.

Despite using the subfields of research study to guide corpus design and collection, the analysis performed in this study was restricted to the main discipline areas. The composition of CAT can be found in Table 2 below.

Table 2. Final composition of the Corpus of Academic Texts (CAT).

Register	Discipline Area	Number of texts	Number of tokens
Articles	Biological Sciences - BIO	200	1,733,310
	Computer and Information Science and Engineering - CISE	200	2,096,007
	Engineering - ENG	200	1,502,550
	Mathematical and Physical Sciences - MPS	200	1,976,715
	Social, Behavioral and Economic Sciences - SBE	200	2,020,657
Subtotal		**1000**	**9,329,239**
Textbooks	Biological Sciences - BIO	30	13,087,559
	Computer and Information Science and Engineering - CISE	30	7,534,488
	Engineering - ENG	30	6,629,986
	Mathematical and Physical Sciences - MPS	30	8,435,077
	Social, Behavioral and Economic Sciences - SBE	30	5,848,916
Subtotal		**150**	**41,536,026**
Total		**1,150**	**50,865,265**

In order to obtain a proof of concept for collocational dimensions as representations of sets of collocates co-occurring across a range of nodes, the methodology described herein was carried out only for the Social, Behavioral, and Economic Sciences subcorpus, characterizing written academic English of that discipline. Therefore, the methodological steps explained in the following sections were performed on the 230 texts (7,869,573 tokens; 709,194 types) comprising the SBE subcorpus.

2.2. Corpus Tagging

All texts in the corpus were tagged for part of speech and lemma with the Tree Tagger[3] (Schmid, 1994). The Tree Tagger was chosen specifically because it provides the

[3] Available for download at http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/.

lemma, “a group of all inflectional forms related to one stem that belong to the same word class” (Kučera & Francis, 1967, p. 1 in Brezina, 2018, p. 40), of each word tagged. As mentioned in the previous chapter, the choice for using lemmas and parts of speech instead of word forms reflects the goal of this study, which is to identify the relationships between collocations, that is, how collocates inter-collocate across nodes, through analysis of semantic preferences and fields. It is thought that those relationships and collocational patterns across texts are not dependent on the varying forms of a word since lexemes make up most lexical collocations (Evert et al., 2017). Homographs were differentiated by part of speech.

A computer script was used to run the Tree Tagger on all texts in the corpus. An example of a tagged text can be seen on Figure 2. The resulting tagged text file is tokenized, with each line of the file containing the original word from the corpus, the part-of-speech (POS) tag[4], and the lemma for that word, separated by tabs. The file is meant to be read vertically; in the case of Figure 2, the first sentence reads: *Respiratory syncytial virus is an exceptional mucosal pathogen*. The first word *Respiratory* is tagged as an adjective (*JJ*) and its lemma is *respiratory*; the verb *is* received the tag for verb in the 3rd person singular present (*VBZ)* and its lemma is *be*.

Figure 2. Example of a text tagged with the Tree Tagger. (corpus file: 1_6_4.txt)

Respiratory	JJ	respiratory
syncytial	JJ	syncytial
virus	NN	virus
is	VBZ	be
an	DT	an
exceptional	JJ	exceptional
mucosal	JJ	mucosal
pathogen	NN	pathogen
.	SENT	.
It	PP	it
specializes	VBZ	specialize
in	IN	in
infection	NN	infection
of	IN	of
the	DT	the
ciliated	JJ	<unknown>
respiratory	JJ	respiratory
epithelium	NN	epithelium

Manual verification of automatic tagging to check for accuracy and corrections were performed as needed. Since the texts that comprise the corpora can be

[4] The Tree Tagger tagset is available in Appendix 1.

considered specialized, there were several discipline-specific words that the tagger could not recognize hence it added the tag <*unknown*> as the lemma of those words. A computer script was used to replace the <*unknown*> tag with the original word from the corpus. As part of the verification of tagging accuracy, 10% of the texts in the SBE corpus were manually checked. During this process, it was noted that four words were consistently tagged incorrectly in those texts: (1) *following*; (2) *learning*; (3) *clustering*; and (4) *bound*. When occurring as *the following* by itself, as in *the following were identified…*, the word *following* was correctly tagged as a noun. However, in every instance when it was followed by a noun, such as *the following criteria*, it was tagged incorrectly as a gerund (VBG) rather than an adjective (JJ). Thus, all those instances were manually changed to reflect the correct tagging (JJ). A further checking of all the other texts in the corpus for this issue revealed the same error, which was promptly corrected. Likewise, instances of *learning* and *clustering* were mistakenly classified as gerund (VBG) rather than as a noun (NN), specifically in the expressions *machine learning, learning algorithms* and *clustering method*. All instances of those expressions were corrected in the corpus. Finally, every instance of *bound* in the expressions *upper bound* and *lower bound* was classified as a verb (VBN) rather than an adjective (JJ) and was also corrected in all texts in which it occurred.

After all corpus files had been tagged and the tags had been fixed as needed, the files were renamed according to the convention: *disciplinecode_registercode_filenumber.txt*. For example, the file *w_academic_articles_sbe_000001.cl1.tagged.fix.txt* was renamed as *5_6_1.txt*. The following codes were used:

a. Discipline:
    a. BIO = 1
    b. CISE = 2
    c. ENG = 3
    d. MPS = 4
    e. SBE = 5
b. Register:
    a. articles = 6
    b. textbooks = 7

2.3. Collocation Identification and Extraction

A number of decisions must be made prior to identification and extraction of collocations, as stated by Evert:

> The resulting set or ranking of collocations depends on many parameters, including the size and composition of the corpus, pre-processing (such as lemmatisation), application of frequency thresholds, the definition of cooccurrence used, and the choice of association measure. It is up to the researcher to find a suitable and meaningful combination of parameters, or to draw on results from multiple parameter settings in order to highlight different aspects of collocativity. (2008, p. 1243)

As mentioned in previous chapters, this study follows an empirical view of collocations, defined as frequently recurring and predictable combinations of words that can be extracted from corpora based on a given window span. In other words, what Sinclair et al. named "significant" collocation, that is, a "regular collocation between two items, such that they co-occur more often than their respective frequencies, and the length of text in which they appear, would predict" (1970/2004, p. 10). Thus, following Sinclair's tradition, collocation is defined here as a word pair that tends to occur in the span of up to four words to either side of the node (Sinclair et al., 1970/2004, p. 13). For the purposes of this study, only complete, lexical collocations were considered for analysis since topical coherence, shared topics, and semantic fields that make up dimensional collocational networks are expressed by lexical words rather than grammatical items. This means that all word pairs extracted for analysis were comprised of lemmatized nouns, adjectives, and verbs. Although adverbs may be classified as lexical words, a choice was made for not including them in this study because they are usually used as adverbials (clause elements) and due to the difficulty in distinguishing adverbs from other word classes, since, as stated by Biber et al., "[a]s clause elements, they border on inserts" (1999, p. 65) rather than lexical words. Moreover, there is a higher frequency of adverbs serving a linking function in academic texts, thereby functioning as adverbials and not lexical words (Biber et al., 1999).

Scripts[5] in Python were written specifically for this study to carry out the tasks related to: (1) identification of window span and node and collocate pairs; (2) identification of the most frequent nodes and their collocates; (3) extraction of word

[5] All Python scripts used in this study were written in collaboration with Dr. Joseph Collentine (Northern Arizona University, Flagstaff, AZ, USA). They can be found in Appendix 2.

pairs belonging to the part of speech of interest (nouns, adjectives, and verbs); (4) calculation of strength of association statistic for nodes and collocates; (5) and extraction of nodes and collocates based on established cut-off criteria. The inspiration for the creating the Python scripts were the scripts developed in Unix Shell by Berber Sardinha's (2017) for his collocational MD analysis study.

As aforementioned, in order to answer the research questions and determine whether collocational dimensions can be seen as representations of sets of collocates co-occurring across a range of nodes, node and collocate pairs were extracted for all written texts, i.e., textbooks and journal articles, of the SBE subcorpus, meant to characterize written academic English of that discipline. After extraction, the node and collocate pairs were then entered into an exploratory factor analysis (See Section 3.4) so as to identify the collocational dimensions of written academic English in Social, Behavioral, and Economic Sciences, followed by a cluster analysis (See Section 3.5) aimed at identifying groupings of nodes, or "node types", in terms of those dimensions.

2.3.1. Identification of Node and Collocate Pairs

For every file in the SBE corpus, the first step was to merge and extract the POS tag and lemma for each word in the corpus. The tag and lemma were joined by an underline ("_") and the tag was shortened to its first letter so that the focus of the analysis was a lemma's primary part-of-speech. For example, *JJ respiratory* became *J_respiratory*. Next, the node and collocate window span was identified for all lemmas of the target part of speech, i.e., all lemmatized adjectives, nouns, and verbs. Therefore, for every lemmatized adjective, noun, and verb (i.e., every potential node), the four words to either side of it were identified (Collocate window: 4L-4R). An example of the resulting database can be found in Figure 3. In the figure, *var1* equals the discipline code, *var2* equals the register, and *var3* equals the file number; the column *word* represents the potential nodes, and the four columns before and after it, the collocate window. End of the sentence boundaries were kept; thus, whenever the four-word window was not full, a 'dummy' word (*none_none*) was added at the point of the sentence boundary and on through the end of that window so that *intra* sentential collocates were considered in the analysis (see lines 2, 3, 15, 16, 17, 18, and 19 in Figure 3).

Figure 3. Sample of node-collocate window span for all lemmatized adjectives, nouns, and verbs as potential nodes in the SBE corpus.

	A	B	C	D	E	F	G	H	I	J	K	L
1	var1	var2	var3	word_minus_4	word_minus_3	word_minus_2	word_minus_1	word	word_plus_1	word_plus_2	word_plus_3	word_plus_4
2	5	6	1	none_none	none_none	WRB_how	DT_the	N_hippocampus	V_preserve	N_order	DT_the	N_role
3	5	6	1	none_none	WRB_how	DT_the	N_hippocampus	V_preserve	N_order	DT_the	N_role	IN_of
4	5	6	1	WRB_how	DT_the	N_hippocampus	V_preserve	N_order	DT_the	N_role	IN_of	N_prediction
5	5	6	1	N_hippocampus	V_preserve	N_order	DT_the	N_role	IN_of	N_prediction	CC_and	N_context
6	5	6	1	N_order	DT_the	N_role	IN_of	N_prediction	CC_and	N_context	V_remember	DT_the
7	5	6	1	N_role	IN_of	N_prediction	CC_and	N_context	V_remember	DT_the	N_sequence	IN_of
8	5	6	1	IN_of	N_prediction	CC_and	N_context	V_remember	DT_the	N_sequence	IN_of	N_event
9	5	6	1	CC_and	N_context	V_remember	DT_the	N_sequence	IN_of	N_event	V_be	J_critical
10	5	6	1	V_remember	DT_the	N_sequence	IN_of	N_event	V_be	J_critical	IN_for	V_derive
11	5	6	1	DT_the	N_sequence	IN_of	N_event	V_be	J_critical	IN_for	V_derive	N_meaning
12	5	6	1	N_sequence	IN_of	N_event	V_be	J_critical	IN_for	V_derive	N_meaning	IN_from
13	5	6	1	N_event	V_be	J_critical	IN_for	V_derive	N_meaning	IN_from	PP$_our	N_experience
14	5	6	1	V_be	J_critical	IN_for	V_derive	N_meaning	IN_from	PP$_our	N_experience	CC_and
15	5	6	1	V_derive	N_meaning	IN_from	PP$_our	N_experience	CC_and	J_guiding	N_behavior	none_none
16	5	6	1	IN_from	PP$_our	N_experience	CC_and	J_guiding	N_behavior	none_none	none_none	none_none
17	5	6	1	PP$_our	N_experience	CC_and	J_guiding	N_behavior	none_none	none_none	none_none	none_none
18	5	6	1	none_none	none_none	none_none	none_none	J_prior	N_investigation	IN_into	DT_the	N_function
19	5	6	1	none_none	none_none	none_none	J_prior	N_investigation	IN_into	DT_the	N_function	IN_of
20	5	6	1	J_prior	N_investigation	IN_into	DT_the	N_function	IN_of	DT_the	J_human	N_hippocampus

A frequency wordlist was then created for all lemmas, per file and for the whole corpus. Next, based on the window span database from the first step, a wordlist of all potential nodes (column *word*) and their frequencies was extracted, as well as a listing of the 2,000 most frequent nodes in the corpus. The following lemmas were disregarded when identifying the 2000 most frequent nodes:

a. Verbs *be*, *do*, and *have*
b. *not*, *ain't*, and *n't*
c. Abbreviation *fig.*
d. Individual letters (for example, *N_p*)

The verbs *be*, *do,* and *have* were disregarded because they play important (modal) grammatical roles, which diminishes their relevance to the identification of lexical collocational patterns. A subset of the database was then created with the window spans for only the 2,000 most frequent nodes identified. It is important to reinforce that only lemmatized adjectives, nouns, and verbs were considered as nodes. Although this is true for the final collocate list as well, all collocates were kept in the window span at this stage, regardless of part of speech, as can be seen in Figure 4. Note that, in contrast with Figure 3 above, the lemmas *N_hippocampus* and *V_be* are no longer considered nodes, since they are not part of the 2,000 most frequent nodes identified in the SBE corpus.

Figure 4. Sample of node-collocate window subset for the 2,000 most frequent nodes in the SBE corpus.

	A	B	C	D	E	F	G	H	I	J	K	L
1	var1	var2	var3	word_minus_4	word_minus_3	word_minus_2	word_minus_1	word	word_plus_1	word_plus_2	word_plus_3	word_plus_4
2	5	6	1	none_none	WRB_how	DT_the	N_hippocampus	V_preserve	N_order	DT_the	N_role	IN_of
3	5	6	1	WRB_how	DT_the	N_hippocampus	V_preserve	N_order	DT_the	N_role	IN_of	N_prediction
4	5	6	1	N_hippocampus	V_preserve	N_order	DT_the	N_role	IN_of	N_prediction	CC_and	N_context
5	5	6	1	N_order	DT_the	N_role	IN_of	N_prediction	CC_and	N_context	V_remember	DT_the
6	5	6	1	N_role	IN_of	N_prediction	CC_and	N_context	V_remember	DT_the	N_sequence	IN_of
7	5	6	1	IN_of	N_prediction	CC_and	N_context	V_remember	DT_the	N_sequence	IN_of	N_event
8	5	6	1	CC_and	N_context	V_remember	DT_the	N_sequence	IN_of	N_event	V_be	J_critical
9	5	6	1	V_remember	DT_the	N_sequence	IN_of	N_event	V_be	J_critical	IN_for	V_derive
10	5	6	1	N_sequence	IN_of	N_event	V_be	J_critical	IN_for	V_derive	N_meaning	IN_from
11	5	6	1	N_event	V_be	J_critical	IN_for	V_derive	N_meaning	IN_from	PP$_our	N_experience
12	5	6	1	V_be	J_critical	IN_for	V_derive	N_meaning	IN_from	PP$_our	N_experience	CC_and
13	5	6	1	V_derive	N_meaning	IN_from	PP$_our	N_experience	CC_and	J_guiding	N_behavior	none_none
14	5	6	1	PP$_our	N_experience	CC_and	J_guiding	N_behavior	none_none	none_none	none_none	none_none
15	5	6	1	none_none	none_none	none_none	none_none	J_prior	N_investigation	IN_into	DT_the	N_function
16	5	6	1	none_none	none_none	none_none	J_prior	N_investigation	IN_into	DT_the	N_function	IN_of
17	5	6	1	J_prior	N_investigation	IN_into	DT_the	N_function	IN_of	DT_the	J_human	N_hippocampus
18	5	6	1	DT_the	N_function	IN_of	DT_the	J_human	N_hippocampus	V_have	V_focus	IN_on
19	5	6	1	DT_the	J_human	N_hippocampus	V_have	V_focus	IN_on	PP$_its	J_more	J_general
20	5	6	1	V_have	V_focus	IN_on	PP$_its	J_more	J_general	N_role	IN_in	J_associative

The next step was to extract the node and collocate pairs and their frequency of occurrence for each file, as well as a unified list for the entire corpus being analyzed. Figure 5 shows an example of the overall listing for the SBE subcorpus. Like in the previous step, at first all collocates were kept, regardless of part-of-speech. The list of collocates was then filtered according to the same criteria as the nodes, that is: only lemmatized adjectives, nouns, and verbs were kept. In addition, the following set of lemmas were removed:

a. Verbs *be*, *do*, and *have*
b. *not*, *ain't*, and *n't*
c. Abbreviation *fig.*
d. Individual letters (for example, *N_p*)
e. Lemmas containing numbers (for example, *N_19th, N_1950s*)

Figure 5. Sample of aggregate node and collocate pairs and their frequency in the SBE corpus.

	A	B	C
1	freq	node	collocate
98	1	J_able	J_acceptable
99	1	J_able	J_accomplished
100	2	J_able	J_accurate
101	2	J_able	J_actual
102	1	J_able	J_adaptive
103	3	J_able	J_additional
104	1	J_able	J_adept
105	2	J_able	J_adequate
106	1	J_able	J_administrative
107	1	J_able	J_affective
108	1	J_able	J_african
109	1	J_able	J_age-related
110	1	J_able	J_agnosic
111	1	J_able	J_alexic
112	2	J_able	J_alternative
113	1	J_able	J_amazing
114	4	J_able	J_ambitious
115	3	J_able	J_american
116	1	J_able	J_amnesic
117	2	J_able	J_anatomical
118	1	J_able	J_ancestral
119	1	J_able	J_ancient
120	2	J_able	J_anthropological
121	1	J_able	J_appropriate

Additionally, the file frequency of the filtered node and collocate pairs was added to the database. As it can be seen in Figure 6, columns were also added to account for the POS tags of nodes, collocates, and node and collocate pairs. Although the analysis was not based on part of speech, the information was kept to facilitate any potential further analyses.

Figure 6. Sample of node and collocate pairs identified in the SBE corpus, showing node and collocate pairs, individual nodes and collocates, their respective POS tags, their joint frequency in the corpus, and the file frequency count.

	A	B	C	D	E	F	G	H
1	node_collocate	node	collocate	node_tag	collocate_tag	node_collocate_tags	freq	file_freq
2	J_able__TAB__J_large	J_able	J_large	J	J	J+J	19	14
3	J_able__TAB__J_many	J_able	J_many	J	J	J+J	23	15
4	J_able__TAB__J_more	J_able	J_more	J	J	J+J	45	21
5	J_able__TAB__J_new	J_able	J_new	J	J	J+J	27	17
6	J_able__TAB__J_other	J_able	J_other	J	J	J+J	28	18
7	J_able__TAB__J_such	J_able	J_such	J	J	J+J	20	16
8	J_able__TAB__N_child	J_able	N_child	J	N	J+N	25	13
9	J_able__TAB__N_firm	J_able	N_firm	J	N	J+N	36	12
10	J_able__TAB__N_group	J_able	N_group	J	N	J+N	22	14
11	J_able__TAB__N_individual	J_able	N_individual	J	N	J+N	23	17
12	J_able__TAB__N_information	J_able	N_information	J	N	J+N	15	13
13	J_able__TAB__N_model	J_able	N_model	J	N	J+N	19	12
14	J_able__TAB__N_other	J_able	N_other	J	N	J+N	15	15
15	J_able__TAB__N_participant	J_able	N_participant	J	N	J+N	14	12
16	J_able__TAB__N_people	J_able	N_people	J	N	J+N	55	21
17	J_able__TAB__N_researcher	J_able	N_researcher	J	N	J+N	29	19
18	J_able__TAB__N_study	J_able	N_study	J	N	J+N	13	12
19	J_able__TAB__V_achieve	J_able	V_achieve	J	V	J+V	16	11
20	J_able__TAB__V_afford	J_able	V_afford	J	V	J+V	35	15
21	J_able__TAB__V_detect	J_able	V_detect	J	V	J+V	21	12
22	J_able__TAB__V_determine	J_able	V_determine	J	V	J+V	17	11
23	J_able__TAB__V_explain	J_able	V_explain	J	V	J+V	22	11
24	J_able__TAB__V_find	J_able	V_find	J	V	J+V	39	23
25	J_able__TAB__V_get	J_able	V_get	J	V	J+V	21	14

The next step was to calculate the strength of association for each word pair in order to extract significant collocations.

2.3.2. Calculation of Strength of Association

A lexical association measure is a way to quantify statistically the attraction between words that co-occur. Different association measures will produce different lists of collocates, emphasizing varying aspects of formulaicity. The literature is rich in the types of association measures that have been proposed, with more than 50 identified (Pecina, 2005). Nevertheless, the “best” association measure for a given study will depend on the research question(s) and what the researcher wishes to highlight in terms of the collocational relationship (Evert, 2008; Brezina, 2018). According to Brezina:

> We can think of most association measures as highlighting collocations along two main dimensions: frequency and exclusivity. Frequency refers to the number of instances in which a node and collocate occur together in a corpus.

> Exclusivity refers to a specific aspect of the collocation relationship where words occur only or predominantly in each other's company. (2018, p. 71)

Most corpus-based studies of collocations have used either Mutual Information (MI) or t-score or a combination of both as the association measure of choice. Both MI and t-score are based on observed and expected frequency of the node and collocate within a given window span. Hunston (2002, p. 71) states that "the MI-score is a measure of the strength of collocation", capturing rare word pairs with low frequencies in the corpus, and the "t-score is a measure of certainty of collocation", capturing the most frequent word combinations. Evert (2005, p. 82-83) and Gablasova, Brezina and McEnery (2017, p. 162) argue that this definition of t-score is not necessarily accurate since it "does not have a very transparent mathematical grounding." According to Hunston (2002) and Gablasova et al. (2017), by not operating on a standardized scale, that is, being dependent on corpus size, the t-score does not allow for direct comparison of collocations in different corpora. In addition, because t-score tends to highlight the most frequent word pairs, the literature has identified a similarity between raw frequency counts and t-score values for collocations (e.g., Durrant & Schmitt, 2009); however, Gablasova et al. stress that "[w]hile all collocations identified by the t-score are frequent, not all frequent word combinations have a high t-score" (2017, p. 163). MI, on the other hand, operates on a logarithmic scale with no theoretical minimum or maximum value, offering a normalized score that enables comparison between different corpora (Hunston, 2002). MI "is negatively linked to frequency in that it rewards lower frequency combinations" (Gablasova et al., 2017, p. 164). This means that the higher the MI value, the more exclusive the association between the two words; thus, technical or specialized terms with lower frequency combinations, which may not be equally distributed or comparable across different corpora, are usually favored. Gablasova et al. also state that caution must be used when interpreting word associations with a high MI value since "the MI-score is not constructed as a (reliable) scale for coherence or semantic unity of word combinations; coherence and semantic unity are an indirect side effect of the measure's focus on rare exclusivity." (2017, p. 163-164)

In this study, the strength of association of each word pair was calculated using the logDice coefficient (Rychly, 2008). Unlike MI and t-score, logDice tends to avoid extreme cases serving as a middle ground between the two association measures

(Berber Sardinha, 2017), as can be seen in Figure 7. In other words, the logDice statistic highlights exclusivity while avoiding the low-frequency bias found in the MI-score (Gablasova et al., 2017) since it “favour[s] collocates which occur exclusively in each other’s company but do not have to be rare” (Brezina, 2018, p. 70). Moreover, the logDice formula does not include expected frequency and is not dependent on corpus size, being a standardized measure on a delimited scale with a fixed theoretical maximum of 14, which makes it preferable to t-score and MI when making comparisons across different corpora. Association measure comparison studies have also identified Dice and LogDice as providing good results for collocation identification (Pecina, 2005; Evert et al., 2017).

Figure 7. Strength of association measures on a frequency and exclusivity scale (adapted from Brezina, 2018, p. 74). LL=Log Likelihood; MI=Mutual Information; MI2=Two-way Mutual Information; MI3=Three-way Mutual Information; MS=Minimum Sensitivity; MU=mu-value c

The logDice formula (Rychly, 2008, p. 9) is

$$14 + \log_2 \frac{2 \times f_{xy}}{f_x + f_y},$$

where f_{xy} is the joint frequency of the node and the collocate within the 4L–4R span, f_x is the frequency of the node in the corpus, and f_y is the frequency of the collocate

in the corpus. As mentioned above, the formula does not include expected frequency nor corpus size, being dependent only on the relative frequencies of the node and collocate pairs in relation to individual frequencies of the node and collocate (Rychly, 2008). According to Rychly (2008), a logDice with value equal to 0 means that there is fewer than 1 co-occurrence of the node and collocate pair, and every 1-point increase in value means that the collocation is twice as frequent. The logDice statistic was calculated, as part of the Python scripts, for the filtered node and collocate pairs comprised of a lemmatized verb, noun, or adjective. Figure 8 illustrates the results of the database with logDice values calculated for target node and collocate pairs. Overall, 17,893 unique collocations were identified in the SBE corpus, based on the collocation extraction criteria described in the next subsection.

Figure 8. Sample of node and collocate pairs identified in the SBE corpus with their respective logDice values.

	A	B	C	D	E	F	G	H	I	J	K
1	node_collocate	node	collocate	node_tag	collocate _tag	node_ collocate_ tags	freq	file_freq	node_corpus _freq	collocate_ corpus_ freq	log_dice
2	J_able__TAB__N_researcher	J_able	N_researcher	J	N	J+N	29	19	2272	3153	7.4526
3	J_able__TAB__V_achieve	J_able	V_achieve	J	V	J+V	16	11	2272	1439	7.1424
4	J_able__TAB__V_afford	J_able	V_afford	J	V	J+V	35	15	2272	401	8.745
5	J_able__TAB__V_detect	J_able	V_detect	J	V	J+V	21	12	2272	525	7.9427
6	J_able__TAB__V_explain	J_able	V_explain	J	V	J+V	22	11	2272	2896	7.124
7	J_able__TAB__V_identify	J_able	V_identify	J	V	J+V	54	29	2272	2685	8.4796
8	J_able__TAB__V_keep	J_able	V_keep	J	V	J+V	21	12	2272	2146	7.2831
9	J_able__TAB__V_meet	J_able	V_meet	J	V	J+V	15	12	2272	1257	7.1218
10	J_able__TAB__V_pay	J_able	V_pay	J	V	J+V	27	13	2272	4305	7.0717
11	J_able__TAB__V_perform	J_able	V_perform	J	V	J+V	17	12	2272	1775	7.1048
12	J_able__TAB__V_understand	J_able	V_understand	J	V	J+V	22	16	2272	3269	7.0235
13	J_absolute__TAB__J_relative	J_absolute	J_relative	J	J	J+J	36	23	437	2225	8.7916
14	J_absolute__TAB__N_term	J_absolute	N_term	J	N	J+N	26	20	437	5316	7.2103
15	J_absolute__TAB__N_value	J_absolute	N_value	J	N	J+N	60	21	437	8641	7.7587
16	J_academic__TAB__N_achievement	J_academic	N_achievement	J	N	J+N	36	15	471	380	10.4369
17	J_accurate__TAB__N_assessment	J_accurate	N_assessment	J	N	J+N	13	11	515	483	8.7375
18	J_accurate__TAB__N_estimate	J_accurate	N_estimate	J	N	J+N	25	13	515	2415	8.1272
19	J_accurate__TAB__N_information	J_accurate	N_information	J	N	J+N	26	17	515	5619	7.1178
20	J_accurate__TAB__N_prediction	J_accurate	N_prediction	J	N	J+N	13	11	515	752	8.3932

2.3.3. Collocation Extraction

The following parameters were used for the extraction of collocations: (1) 2,000 most frequent nodes in the corpus; (2) minimum cut-off value for the strength of association measure; (3) minimum frequency of the collocation in the corpus; and (4) text dispersion, i.e., minimum number of texts in which the collocation must occur. Thus, collocations were extracted only for the 2,000 most frequent nodes in the corpus under

analysis. As mentioned in the previous subsection, the logDice statistic was used to measure the strength of association between the node and the collocate. Only node and collocate pairs with a logDice value of at least 7 were extracted. The minimum frequency of the collocation in the corpus was established as equal or greater to 5, since according to Sinclair, “a language pattern – however defined – has to occur a minimum of twice” (2004, p. 28). Finally, text dispersion—the “distribution of collocates in individual corpus files or corpus parts” (Brezina, 2018, p. 71)—was also added as a parameter for extraction. Text dispersion was considered because “the more evenly dispersed a collocate is in a corpus, the more important it is for the corpus as such” (Brezina, 2018, p. 71). It was established that, in order to be extracted, the node and collocate pair must have occurred in at least 5% of the texts that comprise the corpus being analyzed. Table 3 summarizes the criteria being used in this study to identify and extract collocations:

Table 3. Criteria for collocation identification and extraction.

Criteria	Values
Node + collocate window span	4L–4R
Strength of association measure	logDice
Minimum logDice value	> 7
Minimum collocation frequency	≥ 5
Dispersion	Must occur in > 5% of texts
Filter 1	2,000 most frequent nodes
Filter 2	Only lemmatized adjectives, verbs, and nouns

Thus, nodes and collocates accounting for the 17,893 unique collocations identified in the SBE corpus were extracted based on the criteria described above. The final output of the Python scripts was a dataset containing the nodes as observations (rows) and the collocates as variables (columns). Figure 9 illustrates part of the dataset with nodes and collocates extracted, showing the nodes as rows of observation, the collocates as variables, and the logDice statistic for each node and collocate pair as cell values. A value of 0 indicates the node and collocate pair is not a collocation. In total, the final dataset contained 1,615 nodes and 2,077 collocates. Although there was no dispersion measure accounting for the selection of nodes, i.e., the 2,000 most frequent nodes in the SBE corpus were extracted regardless of in how many texts they appeared, the final dataset contained only 1,615 nodes. This reduction is a result of the cut-offs described herein, which were considered for the

node and collocate pair as a unit. Thus, 315 nodes were eliminated because the node and collocate pairs of which they were a part of either: (1) did not reach the logDice cut-off value; (2) occurred with a frequency of fewer than 5; or occurred in fewer than 5% of the texts in the corpus.

Figure 9. Partial dataset of nodes and collocates extracted from the SBE corpus.

	A	B	C	D	E	F	G	H	I	J	K
1	node	J_social	J_different	J_other	J_new	N_people	J_such	N_rate	J_many	N_researc	N_firm
2	J_ethnic	0	0	0	0	0	0	0	0	0	0
3	N_member	0	7.6519	8.5639	7.9036	0	0	0	0	0	0
4	N_individual	7.52	7.7	7.5414	0	0	0	0	7.2388	0	0
5	N_control	8.8881	0	0	0	0	0	0	0	0	0
6	N_group	9.0744	9.1038	8.9958	0	8.8522	8.0988	0	7.6697	0	0
7	J_different	7.2921	10.3112	7.2178	0	8.7479	0	7.5805	8.8959	0	7.04
8	J_social	8.3564	7.2921	8.0946	7.0899	7.6169	8.2035	0	7.5685	7.8156	0
9	J_other	8.0946	7.2178	7.0665	0	9.3994	8.7806	0	9.3407	7.1948	8.1583
10	J_small	0	0	0	0	0	0	0	7.2299	0	8.4413
11	N_people	7.608	8.7479	9.3975	0	7.1976	7.786	0	9.793	0	0
12	J_large	7.1014	0	0	0	7.203	7.5274	0	0	0	9.0423
13	J_racial	0	0	0	0	0	0	0	0	0	0
14	N_minority	0	0	0	0	0	0	0	0	0	0
15	V_live	0	0	0	0	9.9966	0	0	7.4344	0	0
16	N_difference	0	0	0	0	0	7.0165	7.2057	0	0	0
17	J_particular	0	0	0	0	0	0	0	0	0	0
18	N_membership	0	0	0	0	0	0	0	0	0	0
19	J_such	8.2035	0	8.7806	7.3115	7.786	0	0	7.8688	7.3087	7.4969
20	N_business	0	7.1582	7.1528	8.2013	0	0	0	7.0746	0	7.7713
21	N_age	0	7.3798	0	0	7.0749	0	0	0	0	0
22	N_size	0	7.1389	0	0	0	0	0	0	0	8.258
23	V_form	0	0	0	8.0015	0	0	0	0	0	0
24	V_include	7.7174	0	8.5315	7.0592	7.7098	8.3739	0	8.6373	7.2259	0
25	J_same	0	8.2173	7.2691	0	7.5	0	7.3591	7.1933	0	7.2132

2.4. Exploratory Factor Analysis

As mentioned in Chapter 1, an exploratory multivariate technique, namely exploratory factor analysis, was carried out in order to identify collocational dimensions, that is, the patterns of co-occurrence of collocates across a series of nodes.

Principal Axis Factoring (PAF) has been the extraction method of choice for exploratory factor analysis used in most multidimensional language studies (see Biber, 2014 for a survey of MD studies, Friginal, 2013 and Berber Sardinha & Veirano Pinto, 2014 for recent collections of MD studies, and Berber Sardinha & Veirano Pinto, 2019 for MD methodology)—including the collocational MD analysis developed by Berber Sardinha (2017)—, therefore, it was the preferred extraction method in this study. Principal Components Analysis (PCA) has also been widely used in language studies

(e.g., Diwersy, Evert & Neumann, 2014; Evert & Neumann, 2017; Evert, 2017; Collentine & Collentine, 2018; Biber & Egbert, 2016). According to Evert & Neumann, "both approaches identify latent dimensions based on feature correlations and thus facilitate the visualization of the high-dimensional distribution of a data set" (2017, p. 9), that is, they both attempt to reduce a correlation matrix (*R*-matrix) down to a set of dimensions that capture the most variance in the data. Both PAF and PCA have been shown to provide similar results in language studies, especially with large data sets, and they follow the same general steps of initial extraction, definition of number of factors or components, rotation, and interpretation (Loewen & Gonulal, 2015; Cantos-Gomez, 2019). The main difference lies in that PAF attempts to account for the maximum amount of *common* variance in the smallest number of factors, where "original variables are defined as linear combinations of the factors" in an attempt "to find the presence of latent (unobserved) variables, which explains the pattern of observed variables" (Cantos-Gomez, 2019, p. 106; Loewen & Gonulal, 2015). In contrast, PCA attempts to account for the maximum amount of *total* variance by transforming the variables into linear components, "calculated as linear combinations of the original variables" (Cantos-Gomez, 2019, p. 106; Loewen & Gonulal, 2015). In sum, PCA makes no assumptions about latent variables whereas PAF may identify latent variables in the data being analyzed. Since this study argues that the patterns of collocational variation observed in the original data may be due to variation in topic/subject-matter as a latent (unobserved) variable, the choice of using PAF as an extraction method is deemed justified.

Overall, the steps to perform the factor analysis in this study comprise the following:

1. Definition of factorability of the data.
2. Initial exploratory factor analysis with Principal Axis Factoring (PAF) as the extraction method.
3. Analysis of communalities to identify any variables that may need to be dropped.
4. Definition of optimal number of factors based on analysis of scree plot and eigenvalues.
5. Rotated PAF (with Promax extraction), based on optimal number of factors identified.
6. Analysis of loadings to identify any variables that may need to be dropped.

7. Calculation of factor scores for each node (sum of normalized logDice values of the collocates loading on each factor).
8. Interpretation of each factor and naming of collocational dimensions based on: analysis of semantic preference, semantic field, topic or subject-matter expressed, concordance lines and text excerpts.

All statistical procedures carried out in this study were performed with IBM SPSS Statistics v.23[6].

2.4.1. Factorability of Data

As expected, the dataset from the collocation extraction produced in the last step of the Python scripts (see Section 3.3.3) contained more collocates (variables) than nodes (observations). As aforementioned, for the SBE corpus, the final dataset contained 1,615 nodes and 2,077 collocates. Therefore, in order to meet the observation-to-variable ratio generally recommended to carry out an exploratory multivariate statistic procedure such as factor analysis (Loewen & Gonulal, 2015), the dataset had to be adjusted so that the number of variables could be reduced. It was decided a node-to-collocate ratio of 5:1, where there are 5 observations (nodes) per variable (collocates) (Gorsuch, 2015, p. 350), would meet the sample size assumptions needed to carry out the factor analysis, resulting in a dataset containing 1,615 nodes and 323 collocates. The sample size, i.e., the number of observations (nodes), remained constant and the number of variables (i.e., collocates) was reduced according to the following steps:

1. Calculation of mean logDice value for each collocate (i.e., each column) across all the nodes in the dataset.
2. Ranking of collocates from highest to lowest mean logDice.
3. Calculation of the number of nodes with logDice > 7 for each collocate (i.e., for each column).
4. Ranking of collocates from highest to lowest number of nodes.

[6] https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/product_landing.html

5. If there were any differences in the rankings, conduct a calculation of Spearman's and Pearson correlations to compare them.
6. Determination of which ranking should be used to select the collocates for the final dataset.
7. Selection of collocates according to the ranking chosen.

The two rankings, mean logDice for each collocate and number of nodes per collocate, were chosen as a principled way to reduce the number of variables in the dataset. It was expected that both rankings would be similar, given the nature of the data and Spearman's and Pearson correlations were conducted to verify that assumption. Figures 10 through 13 illustrate this procedure. Figure 10 shows a sample of the calculation of the mean logDice and the count of nodes per collocate (lines 1618 and 1619). The calculation of the mean logDice was done with the formula for AVERAGE on Excel and the calculation of number of nodes per collocate was done with the formula COUNTIF. For example, for the collocate in column B (*J_able*), the formula for the mean logDice value was AVERAGE(B2:B1616), giving a mean of 0.051, and the formula for the count of how many nodes *J_able* collocates with was COUNTIF(B2:B1616, ">7"), showing that it collocates with 11 different nodes. The formulas were applied to each column. Next, the collocates were ranked from highest to lowest for both calculations, as the example in Figure 11 shows.

Figure 10. Sample of the calculation of mean logDice and number of nodes per collocate for the SBE corpus node and collocate pairs dataset.

	A	B	C	D	E	F	G	H	I
1	node	J_able	J_absolute	J_abundan	J_academic	J_accountable	J_accurate	J_active	J_actual
1600	V_train	0	0	0	0	0	0	0	0
1601	V_travel	0	0	0	0	0	0	0	0
1602	V_treat	0	0	0	0	0	0	0	0
1603	V_try	0	0	0	0	0	0	0	0
1604	V_turn	0	0	0	0	0	0	0	0
1605	V_underlie	0	0	0	0	0	0	0	0
1606	V_understand	7.0235	0	0	0	0	0	0	0
1607	V_use	0	0	0	0	0	0	0	0
1608	V_vary	0	0	0	0	0	0	0	0
1609	V_wait	0	0	0	0	0	0	0	0
1610	V_want	0	0	0	0	0	0	0	0
1611	V_watch	0	0	0	0	0	0	0	0
1612	V_wear	0	0	0	0	0	0	0	0
1613	V_win	0	0	0	0	0	0	0	0
1614	V_work	0	0	0	0	0	0	0	0
1615	V_write	0	0	0	0	0	0	0	0
1616	V_yield	0	0	0	0	0	0	0	0
1617									
1618	Mean logDice	0.051078	0.014712446	0.004471	0.006462477	0.005385882	0.020046873	0.009734	0.022371
1619	Node Count	11	3	1	1	1	4	2	5

Figure 11. Sample of both collocate rankings, showing collocates ranked from highest to lowest according to the count of nodes per collocate.

	A	B	C	D	E
1	**Collocate**	**Mean logDice**	**Rank_Mean**	**Count of Nodes**	**Rank_Nodes**
2	J_social	0.54779226	1	113	1
3	J_different	0.491529536	2	104	2
4	J_other	0.428625882	4	90	3
5	J_new	0.422689969	5	89	4
6	N_people	0.414350898	6	87	5
7	J_such	0.402095356	10	87	6
8	N_rate	0.432173932	3	86	7
9	J_many	0.407161238	7	86	8
10	N_research	0.404671393	8	84	9
11	N_firm	0.404173065	9	84	10
12	J_high	0.401086997	11	82	11
13	N_effect	0.374658885	14	78	12
14	J_human	0.378174613	12	77	13
15	V_use	0.369774489	15	77	14
16	N_group	0.365399381	16	76	15
17	N_study	0.362727245	17	75	16
18	V_make	0.362425697	18	75	17
19	N_price	0.374781734	13	74	18
20	V_increase	0.347140681	20	73	19
21	N_change	0.347994303	19	72	20

Although very similar, the rankings did show a few differences (e.g., *N_rate* collocates with 86 nodes, ranking number 7 in the node count ranking, but it has a high

mean logDice value, appearing in number 3 in that raking). Hence, Spearman's and Pearson correlations were calculated to verify if the rankings were significantly correlated and to support the decision on which ranking to use for the final selection of collocates. Both correlations had a value of 1.000 (*p*=0.000) (Figure 12).

Figure 12. Results of the Pearson and Spearman's correlations for the two sets of rankings: mean logDice and node count per collocate.

Pearson Correlation

		Rank_Mean	Rank_NodeCount
Rank_Mean	Pearson Correlation	1	1.000**
	Sig. (2-tailed)		0.000
	N	2077	2077
Rank_NodeCount	Pearson Correlation	1.000**	1
	Sig. (2-tailed)	0.000	
	N	2077	2077

**. Correlation is significant at the 0.01 level (2-tailed).

Spearman's rho correlation

			Rank_Mean	Rank_NodeCount
Spearman's rho	Rank_Mean	Correlation Coefficient	1.000	1.000**
		Sig. (2-tailed)		0.000
		N	2077	2077
	Rank_NodeCount	Correlation Coefficient	1.000**	1.000
		Sig. (2-tailed)	0.000	
		N	2077	2077

**. Correlation is significant at the 0.01 level (2-tailed).

Since one of the principles being argued for collocational dimensions is that collocates tend to co-occur across a series of nodes forming different collocational patterns, a decision was made to use the ranking of number of nodes per collocate to determine the final list of collocates for the factor analysis. Therefore, the 323 collocates with the highest number of nodes were selected to comprise the final dataset. A script in Python was used to reduce the original output to the final 1,615 nodes and 323 collocates. The final list of 1,615 nodes and 323 collocates can be found in Appendix 3. The dataset was then uploaded into SPSS v.23 so that the multivariate analysis could be performed.

Next, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity were calculated. As it can be seen in Figure 13, the KMO value was .775, indicating a "good" (Field 2009) or almost "meritorious" (Kaiser 1974) sampling adequacy, and the Bartlett's test indicates that the null hypothesis of no correlations existing among variables can be rejected (*p*<.001). In addition, the

correlation matrix (*R*-matrix) showed no correlations above .90, indicating a lack of multicollinearity in the dataset. Having established the adequacy of the dataset, the exploratory factor analysis was carried out, whose procedure is detailed in the next subsection.

Figure 13. Results of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.775
Bartlett's Test of Sphericity	Approx. Chi-Square	219011.262
	df	52003
	Sig.	0.000

2.4.2. Exploratory Factor Analysis

2.4.2.1. Initial Factor Analysis

As aforementioned, an initial exploratory factor analysis was carried out with Principal Axis Factoring as the extraction method, following the path *Analyze* > *Dimension Reduction* > *Factor Analysis* on SPSS v.23. All of the variables were entered into the factor analysis (Figure 14). As shown on the top right corner of Figure 14, a number of configurations can be made to carry out the factor analysis (*Descriptives, Extraction, Rotation, Scores,* and *Options*). For the initial factor analysis, only *Descriptives, Extraction, and Rotation* were configured, as described below.

Figure 14. SPSS v.23 – Factor Analysis: Variable selection.

The statistics for determining the factorability of the data are chosen in *Descriptives* (Figure 15). The results of the statistics chosen in this step were discussed in the previous subsection. An initial solution, the coefficients, determinant, and KMO and Bartlett's test of sphericity are selected. The *Univariate descriptives* provides means and standard deviations of each variable. The *R*-matrix, i.e., the correlation matrix, is provided by choosing *coefficients*. The *Determinant* tests for multicollinearity in the data. The KMO and Bartlett's test of sphericity are used to determine the data's sampling adequacy.

Figure 15. SPSS v.23 – Factor Analysis: Descriptives.

The extraction method is chosen in the *Extraction* dialog box, as shown in Figure 16. As previously explained, Principal Axis Factoring was used as the extraction method in this study. Since this is the initial extraction, an unrotated factor solution is selected, as well as the scree plot, based on eigenvalues greater than 1. The number of iterations for convergence was set at 200, given the large amount of observations in the data set. The next configuration is *Rotation*. The initial solution is unrotated (Rotation method *None*) so that the maximum number of factors may be extracted (Figure 17). The initial factor analysis is then carried out.

Figure 16. SPSS v.23 – Factor Analysis: Extraction.

Figure 17. SPSS v.23 – Factor Analysis: Rotation.

Besides the statistics for determining factorability of the data set, the unrotated factor analysis with PAF as the extraction method produces a table of initial and extraction communalities, the total variance explained in the data, a scree plot, and a matrix of all factors with eigenvalues greater than 1. Communalities show estimates of the amount of variance in each variable that is accounted for by all factors (initial communalities) and by the factors in the unrotated factor solution (extraction communalities) (Cantos-Gomez, 2013). According to Loewen & Gonulal, "[h]igh communalities are desired because they indicate that the EFA results perform well in accounting for variance within the variables" (2015, p. 190). There were no initial communalities below .30 and only one variable had a communality below .30 (*N_attention* = 0.213) in the extraction, hence none of the variables were dropped from the data set. A sample of the communalities table can be found in Figure 18.

Figure 18. SPSS v.23 – Factor Analysis – Communalities (sample). (Extraction method: Principal Axis Factoring)

Communalities

	Initial	Extraction
J_large	.733	.807
V_shape	.726	.785
J_such	.725	.784
N_kind	.742	.761
J_different	.740	.757
N_aspect	.769	.755
J_consistent	.726	.737
N_woman	.704	.733
V_suggest	.721	.731
J_other	.733	.730
N_characteristic	.653	.723
N_product	.668	.716
J_particular	.666	.713

In a factor analysis, each factor accounts for the maximum variance between the variables, where the first factor will always account for the most variance across all variables, the second will account for variance in the variables remaining after the first factor is extracted, the third, in the variables remaining after the second factor is extracted, and so forth, until 100% of the variance has been accounted for (cf. Biber, 1988). The proportion of variance initially found in each factor is explained by eigenvalues. Eigenvalues "are direct indices of the amount of variance accounted for by each factor" (Biber, 1988, p. 82). As the name implies, the total variance explained table (Figure 19) brings the initial eigenvalues, the percentage of variance captured by each factor, as well as the cumulative variance, and the extraction sums of squared loadings for factors with eigenvalues greater than 1. Initial eigenvalues are found in the column *Total*. The ratio of the amount of variance captured by each factor to the total variance across all variables is represented as a percentage in the *% of Variance* column. The cumulative percentage of variance is given in the *Cummulative %* column. The extraction sums of squared loadings part of the table shows the same information, but only for the factors with eigenvalues greater than 1 which were extracted. The variance is now spread only across the number of factors extracted rather than all factors initially extracted. In addition to the total variance table, eigenvalues are also shown visually in the form of a scree plot (Figure 20).

Figure 19. SPSS v.23 – Factor Analysis – Total variance explained (sample). (Extraction method: Principal Axis Factoring)

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	18.419	5.702	5.702	18.012	5.576	5.576
2	9.455	2.927	8.630	9.027	2.795	8.371
3	7.241	2.242	10.872	6.822	2.112	10.483
4	6.375	1.974	12.845	5.945	1.841	12.324
5	5.386	1.668	14.513	4.939	1.529	13.853
6	5.066	1.569	16.081	4.625	1.432	15.285
7	4.505	1.395	17.476	4.071	1.260	16.545
8	4.027	1.247	18.723	3.589	1.111	17.656
9	3.876	1.200	19.923	3.425	1.060	18.717
10	3.579	1.108	21.031	3.158	.978	19.694
11	3.521	1.090	22.121	3.072	.951	20.645
12	3.203	.992	23.113	2.761	.855	21.500
13	3.164	.979	24.092	2.730	.845	22.345
14	3.030	.938	25.030	2.561	.793	23.138
15	2.950	.913	25.943	2.510	.777	23.915
16	2.839	.879	26.822	2.411	.746	24.662
17	2.791	.864	27.687	2.346	.726	25.388
18	2.670	.827	28.513	2.222	.688	26.076
19	2.636	.816	29.329	2.180	.675	26.751
20	2.622	.812	30.141	2.168	.671	27.422

Figure 20. SPSS v.23 – Factor Analysis – Scree plot.

Since the goal of using a factor analysis technique is to reduce the number of variables into interpretable factors with minimum loss of information, it is necessary to identify the number of factors that should be retained—i.e., extracted and rotated—in a final solution, such that the retained factors are fewer than the number of variables. Determining the correct number of factors to extract is crucial since it can significantly affect the interpretation of the results (Loewen & Gonulal, 2015). There are a number of ways to determine how many factors should be extracted. Kaiser's criterion (1960) of retaining factors with eigenvalues greater than 1 is the default in SPSS v.23, and it is used in the initial extraction, as mentioned above. Since it is the default choice, it is commonly used, but it is considered most accurate "when there are fewer than 40 variables and the sample size is adequate" (Gorsuch, 1983, 1990 in Loewen & Gonulal, 2015, p. 194). In this study, there were 100 factors with eigenvalues greater than 1, accounting for 68% of the variance in the data. Therefore, an analysis of the scree plot and total variance table was conducted to determine the number of factors

to be extracted. The scree plot is a graphic representation of the eigenvalues, shown in descending order, and it is commonly used in MD studies for this purpose (Cantos-Gomez, 2019). According to Loewen & Gonulal, “the cutoff point for selecting factors is the point of inflexion, which is the sharp descent, or elbow, in the slope of the plot” (2015, p. 196), that is, “where the sharp drop in eigenvalues ends and the curve of the plot starts levelling off” (Brezina, 2018, p. 166). The number of factors to the left of the point of inflexion is then extracted. Points of inflexion were identified at Factors 5, 8 and 12, which meant extracting either 4, 7, or 11 factors, representing 12.8%, 17.5%, and 22.1% of variance, respectively. Therefore, 4-, 7-, and 11-factor extractions were performed as rotated factor analyses, with the 11-factor solution being considered the most interpretable. As shown in the total variance explained table, the percentages of variance that each factor is contributing are considered low, showing the variance to be largely spread out. This may be because there are nodes with sets of collocates similar to those of other nodes. In addition, the variables (collocates) and the observations (nodes) are linguistically the same: nouns, verbs, and adjectives. In addition, low percentages of variance are not uncommon in exploratory factor analyses with datasets containing lexical items as variables as well as a large number of zeroes, as is the case here and in previously reported lexical and collocational MD analyses (Berber Sardinha, 2017; Berber Sardinha et al., forthcoming; Delfino et al., 2018). The methodology for the rotated factor analysis is described in the next subsection.

2.4.2.2. Rotated Factor Analysis

As mentioned above, none of the variables were dropped based on the analysis of communalities. All variables were then included in the rotated factor analysis. The goal of performing a rotated factor analysis is to simplify the factorial solution. The procedure follows the same steps as the initial factor analysis, with the exception of the number of factors to be extracted and the rotation method. Figure 21 shows the configuration in the *Extraction* dialog box in SPSS v.23. Both options under *display* are now unchecked and the factors are to be extracted based on a fixed number of factors. The number of factors to be extracted is inserted into box provided.

Figure 21. SPSS v.23 – Factor Analysis: Extraction – rotated.

The rotation method is chosen in the *Rotation* dialog box (Figure 22). Rotating the factors allows for a simpler data structure where "each factor is characterized by the few features that are most representative of a particular amount of shared variance" (Biber, 1988, p. 84). It is important to highlight that rotation of the factors does not alter the total variance found in the data, rather it alters how the total variance is distributed among the factors (Cantos-Gomez, 2013). Because this study deals with language data, it can be assumed that factors will be somewhat related, therefore, Promax was chosen as the rotation method (oblique rotation), which allows factors to correlate. The default Kappa value was kept.

Figure 22. SPSS v.23 – Factor Analysis: Rotation – rotated.

The default cut-off for factor loadings, "which indicate the strength of the association between each variable and each factor" (Loewen & Gonulal, 2015, p. 199) is .1 in SPSS v.23. However, it varies in the MD analysis literature from .10 to .40 (Cantos-Gomez, 2013, 2019; Brezina, 2018). In this study, the .20 cut-off value was used. Thus, variables with factor loadings lower than .20 were not considered in the interpretation of the factor. It is assumed that variables with loadings lower than .20 are not contributing as much and do not influence the factor. It is possible to suppress loadings with absolute values below .20 in the *Options* dialog box (Figure 23). The option *sorted by size* was also checked so that the variables in the factors would appear in ascending order of loading magnitude.

Figure 23. SPSS v.23 – Factor Analysis: Options.

Since there were no changes in the variables entered into the rotated factor analysis, the communalities and cumulative variance remained the same. The only alteration was in the Rotation Sums of Squared Loadings values, which now are shown only for the 11 factors extracted (Figure 24).

Figure 24. SPSS v.23 – Factor Analysis – Total variance explained (sample) – rotated. (Extraction method: Principal Axis Factoring)

Total Variance Explained

Factor	Initial Eigenvalues			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total
1	18.419	5.702	5.702	10.914
2	9.455	2.927	8.630	10.554
3	7.241	2.242	10.872	6.115
4	6.375	1.974	12.845	8.569
5	5.386	1.668	14.513	8.053
6	5.066	1.569	16.081	9.726
7	4.505	1.395	17.476	6.233
8	4.027	1.247	18.723	6.983
9	3.876	1.200	19.923	6.589
10	3.579	1.108	21.031	5.212
11	3.521	1.090	22.121	4.398
12	3.203	.992	23.113	

Oblique rotations such as Promax include two tables as part of the results: a pattern matrix and a structure matrix. The difference between them lies in the factor loadings displayed. The structure matrix takes into account relationships between factors whereas the unique contribution of the variable to a factor is what determines the factor loadings in the pattern matrix (Loewen & Gonulal, 2015). In language studies, the pattern matrix is traditionally used for interpretation since it "accentuates the differences among the factors and is therefore often more meaningful and interpretable" (Loewen & Gonulal, 2015, p. 197). The pattern matrix shows the factor loading (positive or negative) of each variable in each of the factors extracted, where factor loadings are considered "measures of importance of the individual variables in relation to a particular factor" (Brezina, 2018, p. 167). Figure 25 illustrates a sample of the pattern matrix table for the 11-factor solution. The full table is presented in the results chapter. Variables are kept in the factor in which they have the highest absolute loading. Note that since the option to suppress loadings with absolute values below .20, the matrix already shows which variables loaded with the highest loading in a factor. If a variable has a loading higher than .20 in more than one factor, such as it is shown for *N_wage*, *N_percent*, and *N_share* in the figure, the highest absolute loading determines to which factor the variable belongs. In total, considering the .20 cut-off, 286 variables loaded into the 11 factors, a drop of 37 variables from the original set. Hence, those three variables shown were included in factor 1. The factor loadings were then used to calculate factor scores for each observation in the data.

Figure 25. SPSS v.23 – Factor Analysis – Pattern Matrix (sample). (Extraction method: Principal Axis Factoring. / Rotation method: Rotation Method: Promax with Kaiser Normalization.)

Pattern Matrix[a]

	Factor										
	1	2	3	4	5	6	7	8	9	10	11
N_wage	.574							.208			
J_low	.560										
N_price	.555										
N_rate	.554										
V_increase	.549										
J_average	.528										
J_high	.518										
N_increase	.518										
N_income	.491										
N_investment	.489										
N_capital	.481										
N_cost	.480										
N_level	.478										
N_return	.469										
N_percent	.467							.240			
V_reduce	.457										
N_profit	.456										
N_share	.452					.228					
V_expect	.450										
N_firm	.442										
J_total	.442										
N_risk	.436										
N_tax	.423										
N_production	.420										

2.4.3. Factor Scores

A factor score is traditionally used in MD analyses to characterize a text (usual unit of observation) in terms of each of the factors identified. Mean factor (or dimension) scores are then used to place registers along the dimensions. In this study, factor scores were computed for each node, which is the unit of observation, by adding up the standardized logDice values for the collocates loading in each factor. Therefore, each node received a score for each of the factors, representing the incidence of the group of co-occurring collocates for that node in a given factor.

There are different ways of computing factor scores. SPSS v.23 brings three different statistical methods for obtaining factor scores: regression, Bartlett, and Anderson-Rubin. The three methods produce similar results and factor scores are computed from all variable values across the factors. This means that the values of all variables are taken into account in the calculation for each factor, regardless of whether the variable loaded into the factor. A simpler method, and the most commonly

used in MD studies, consists of summing up the standardized values of only the variables that loaded in each factor. Summing up those values means that the fact that variables load on different factors (with varying values) is not taken into consideration. The factor scores for a given factor are comprised of only the standardized values for the variables in that factor and the values of all other variables are omitted.

In this study, factor scores were computed as the sum of the standardized logDice values for the variables loading in each factor. The first step in computing factor scores this way consists of standardizing the data, that is, scaling the logDice values as standard deviation units. The process comprises standardizing the logDice values to a mean of 0 and a standard deviation of +1 or -1. This can be accomplished by converting the values into Z-scores. In SPSS v.23, this is done by following the path *Analyze* > *Descriptive Statistics* > *Descriptives.* In the dialog box (Figure 26), all variables are selected, and the box *save standardized values as variables* is checked.

Figure 26. SPSS v.23 – Descriptive Statistics – Standardization.

The new standardized variables are added to the data set and are marked with a *Z* in front of their original name; for example, the standardized variable created from *N_wage* is called *ZN_wage*. Using the newly created standardized variables, factor scores are then computed with a numeric expression. In SPSS v.23, this is done with the *compute variable* command, following the path *Transform* > *Compute Variable*. In the dialog box, the new variable name in inserted in the box for *Target variable* and

the standardized variables are added to the *numeric expression* box, as shown in Figure 27. The figure shows the computation of factor scores for Factor 9. The name of the new variable is *F9*. The numeric expression consists of the sum of the standardized variables with positive loadings in that factor minus the sum of the standardized variables with negative loadings in that factor. Thus, for example, the numeric expression for Factor 9 is as follows:

Factor 9 = (ZV_learn + ZV_use + ZV_develop + ZJ_new + ZV_allow + ZV_give + ZV_require + ZV_need + ZJ_available + ZV_apply + ZV_get + ZV_share + ZN_information + ZV_make + ZN_use + ZV_produce + ZJ_common + ZV_take) - (ZN_impact)

This procedure was carried out for the 11 factors extracted. The numeric expression for each factor can be found in Appendix 4.

Figure 27. SPSS v.23 – Factor Analysis – Computing factor scores.

2.4.4. Interpretation and naming of factors

Once the factor analysis procedures were performed, the next step comprised the interpretation of each of the 11 factors extracted and their naming. In order to support interpretation, a script in Python was written to extract the collocations that comprised the collocates that loaded in each factor as well as the frequency of node and collocates in the collocations extracted. An example of the output can be found in Figure 28, which shows a sample of the collocations made up by collocates that loaded in factor 1.

Figure 28. Sample of collocations in the SBE corpus comprised of collocates from Factor 1.

	A	B	C	D	E	F	G	H
1	node_collocate	node	collocate		node_factor	count_node	collocate_factor	count_collocate
2	J_able__TAB__V_pay	J_able	V_pay		V_increase	31	N_rate	86
3	J_absolute__TAB__J_relative	J_absolute	J_relative		J_high	30	N_firm	84
4	J_absolute__TAB__N_value	J_absolute	N_value		N_increase	27	J_high	82
5	J_additional__TAB__N_cost	J_additional	N_cost		N_price	26	N_price	74
6	J_additional__TAB__V_raise	J_additional	V_raise		N_rate	24	V_increase	73
7	J_aggregate__TAB__N_level	J_aggregate	N_level		J_low	24	N_value	69
8	J_agricultural__TAB__N_labor	J_agricultural	N_labor		N_firm	23	N_market	69
9	J_agricultural__TAB__N_production	J_agricultural	N_production		V_reduce	22	N_level	63
10	J_annual__TAB__J_average	J_annual	J_average		N_cost	20	N_increase	57
11	J_annual__TAB__N_growth	J_annual	N_growth		N_level	18	N_cost	57
12	J_annual__TAB__N_income	J_annual	N_income		N_wage	18	V_reduce	53
13	J_annual__TAB__N_increase	J_annual	N_increase		J_average	18	N_number	53
14	J_annual__TAB__N_percent	J_annual	N_percent		N_market	17	N_income	50
15	J_annual__TAB__N_rate	J_annual	N_rate		J_total	17	N_amount	49
16	J_available__TAB__N_amount	J_available	N_amount		N_income	17	J_low	48
17	J_average__TAB__J_average	J_average	J_average		N_percent	16	N_size	45
18	J_average__TAB__J_high	J_average	J_high		N_value	15	N_percent	45
19	J_average__TAB__J_less	J_average	J_less		N_risk	15	V_pay	42
20	J_average__TAB__J_low	J_average	J_low		V_pay	15	N_worker	40
21	J_average__TAB__N_amount	J_average	N_amount		N_labor	15	N_money	40
22	J_average__TAB__N_firm	J_average	N_firm		N_capital	15	N_labor	39
23	J_average__TAB__N_growth	J_average	N_growth		N_investment	14	V_lead	37
24	J_average__TAB__N_income	J_average	N_income		N_tax	14	N_risk	37
25	J_average__TAB__N_increase	J_average	N_increase		J_large	13	N_wage	36
26	J_average__TAB__N_level	J_average	N_level		N_production	13	N_production	34
27	J_average__TAB__N_number	J_average	N_number		N_worker	13	N_interest	33

As discussed in Chapter 1, the interpretation of the collocates and collocations that comprise each factor was based on analysis of: (1) the co-occurring collocates in the factor; (2) individual collocates with respect to particular node words; and (3) the distribution of the sets of collocates across groupings of nodes. It also took into consideration semantic preference (Stubbs, 2007), lexical sets (Sinclair & Jones, 1974-96) and semantic fields (Brinton & Brinton, 2010), aboutness (Halliday & Hasan,

1976; Phillips, 1985, 1989), topics (Berber Sardinha, 1997), and subject-matter (Schütze, 1998). At the level of the text, lexical or semantic field is defined as "a segment of reality symbolized by a set of related words [which] share a common semantic property. Most often, fields are defined by subject matter" (Brinton & Brinton, 2010, p. 112). In this sense, a lexical or semantic field may be defined by the set of collocations that co-occur in the corpus, which in turn contribute to build topical coherence in discourse. At the level of the word pairs, "semantic preference refers to what has traditionally been known as a lexical field: a class of words which share some semantic feature" (Stubbs, 2007, p. 178). Semantic preference, "the restriction of regular co-occurrence of items which share a semantic feature" (Sinclair, 1998, p. 16), is related to the topic of the surrounding co-text. In other words, the co-occurrence of lexical sets of semantically related words, may reflect and typify topics that are characterized by the relationships underlying the variation in collocational usage.

The analysis of the distribution of the sets of collocates across groupings of nodes as well as the characterization of the role those extended collocational patterns play in weaving discourse lexically was enabled by carrying out a hierarchical cluster analysis for the factors identified in the exploratory factor analysis, as described in the next section, as well as by conducting descriptive statistics of the distribution of nodes across texts and discipline subfields with respect to the dimensions.

2.5. Cluster Analysis

As aforementioned, a cluster analysis was conducted in order to identify "node types" or categories according to which nodes can be grouped in relation to the collocational patterns identified by the dimensional collocational analysis. A cluster analysis provides an exploratory, bottom-up way to reduce observations and identify underlying groups that share similar characteristics while, at the same time, separating those that are different (Staples & Biber, 2015). In this study, a hierarchical cluster analysis (HCA) was conducted in SPSS v.23, as detailed below, with the input being the factor scores for the nodes in each of the 11 factors identified in the factor analysis. The hierarchical cluster analysis allows the researcher to visualize possible cluster solutions and it supports the selection of number of clusters for subsequent analysis.

2.5.1. Assumptions

Before performing the cluster analysis, it is necessary to check for absence of multicollinearity or singularity in the data. This was accomplished with two different statistical procedures: bivariate correlations and linear regression. Bivariate correlations (Pearson correlations) between the independent variables, or in this case, the factors, were run following the path *Analyze* > *Correlate* > *Bivariate*, as shown in Figure 29. The correlation matrix showed no bivariate correlations above +/- .90. The highest correlation value was .465, between factors 6 and 8.

Figure 29. SPSS v.23 – Cluster Analysis – Bivariate correlations.

Then, tolerance and VIF (Variance Inflation Factor) values were checked using linear regression. Presence of multicollinearity is evidenced by tolerance values below .40 and VIF values above 2.50. This was checked following the path *Analyze* >

Regression > *Linear*. Factor 1 was chosen as the dependent variable and the remaining 10 factors were added as independent variables. Then, in the *Statistics* dialogue box, only *Collinearity Diagnostics* was checked (Figure 30).

Figure 30. SPSS v.23 – Cluster Analysis – Linear Regression.

As it can be seen in Figure 31, the collinearity statistics are all within the recommended range, showing no tolerance values below .40 and no VIF values above 2.50. Hence, the assumption of no multicollinearity or singularity is met.

Figure 31. SPSS v.23 – Cluster Analysis – Collinearity diagnostics.

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Factor2	.704	1.421
	Factor3	.683	1.463
	Factor4	.699	1.431
	Factor5	.769	1.301
	Factor6	.636	1.573
	Factor7	.768	1.303
	Factor8	.717	1.394
	Factor9	.737	1.357
	Factor10	.879	1.138
	Factor11	.885	1.130

a. Dependent Variable: Factor1

2.5.2. Initial Hierarchical Cluster Analysis (HCA)

Since the assumptions were met, an agglomerative hierarchical cluster analysis (HCA) was performed following the path *Analyze* > *Classify* > *Hierarchical Cluster Analysis*. The variables were the 11 factors from the factor analysis and the cases (or observations) were the nodes (Figure 32).

Figure 32. SPSS v.23 – Hierarchical Cluster Analysis.

The default options were kept in the *Statistics* dialogue box, with only *Agglomeration schedule* checked. The *Agglomeration schedule* provides the distance coefficients that are used to analyze the cluster solutions and inform the choice of the optimal number of clusters. Under *Plots*, the *Dendrogram* graph was selected and *None* was checked for *Icicle* plots. See Figure 33 for the configurations used in both dialogue boxes.

Figure 33. SPSS v.23 – Hierarchical Cluster Analysis – Statistics and Plots.

The cluster method should be chosen in the *Method* dialogue box (Figure 34). There are several options and the Ward's Method was chosen. According to Staples & Biber (2015), the Ward's method is the most frequently used method for hierarchical cluster analyses and it "measures dissimilarity between clusters in relation to the 'loss of information' or increase in the error sum of squares by joining two clusters" (Aldenderfer & Blashfield, 1984, in Staples & Biber, 2015, p. 252). The *squared Euclidean distance* should be chosen with the Ward's method as the measure. SPSS v.23 provides a number of options to transform values, z-scores being the most common (Staples & Biber, 2015). Since the calculation of factor scores were already based on standardized values, using z-scores, the option *None* was chosen.

Figure 34. SPSS v.23 – Hierarchical Cluster Analysis – Method.

The output of the initial HCA provides two main pieces of information that are used to interpret the cluster formation and to inform the decision of the optimal number of clusters: the dendrogram plot and the agglomeration schedule table. The dendrogram plot allows for visualization of the hierarchical clusters formed based on the Ward's method and measures chosen. According to Staples & Biber, objects are joined successively into larger clusters: "[a]t first all of the cases are individual clusters. The two most similar clusters are then fused and distances recalculated. More and more objects are linked together until the last step, when all objects are joined" (2015, p. 255). The dendrogram (seen in Figure 35) shows a few possible, clear cluster solution, namely: 2-, 3-, or 4-cluster solutions. However, it is not conclusive enough by itself to determine the optimal number of clusters.

Figure 35. SPSS v.23 – Hierarchical Cluster Analysis – Dendrogram for 1,615 cases (nodes).

Thus, the agglomeration schedule table provided as part of the SPSS v.23 output should also be analyzed. A sample of the table can be seen in Figure 36, sorted by the coefficients, from largest to smallest. In particular, an analysis of the "fusion coefficients, or the numerical value at which cases merge to form a cluster" (Aldenderfer & Blashfield, 1984, in Staples & Biber, 2015, p. 255), provides a more quantitative way to determine the optimal number of clusters. This is accomplished by analyzing the coefficients column, "which indicates the within-cluster sum of squares at the point at which the last two clusters were joined" (Staples & Biber, 2015, p. 257). According to Aldenderfer and Blashfield, the optimal number of clusters is indicated by the point in which "the difference between the fusion coefficients starts to flatten out, which indicates that no new information is being added with the addition of new clusters" (1984 in Staples & Biber, 2015, p. 257). Therefore, the distance between each fusion coefficient and the one immediately before it must be calculated. The differences between the coefficients can then be plotted in relation to the number of clusters and the point where it starts to flatten out can be clearly visualized. This was done using Microsoft Excel. The agglomeration table was sorted according to the coefficients, from largest to lowest. Then a column was added and labeled 'distance'. The distance was calculated as the subtraction of the last coefficient value from the next coefficient. For example, as shown in Figure 36, the last clustering coefficient value was 2231918.155 and the next coefficient value was 1836373.519. Thus, the distance between them was computed as 395544.6 (i.e., 2231918.155 -

1836373.519). The distance was computed for all coefficient values and it was then plotted using a line graph, which can be seen in Figure 37.

Figure 36. SPSS v.23 – Hierarchical Cluster Analysis – Agglomeration schedule (sample).

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1614	1	30	2231918.155	1613	1595	0
1613	1	84	1836373.519	1612	1599	1614
1612	1	44	1548322.257	1611	1610	1613
1611	1	11	1396296.821	1609	1607	1612
1610	44	63	1301057.895	1603	1608	1612
1609	1	2	1208220.582	1606	1601	1611
1608	63	65	1131972.284	1564	1594	1610
1607	11	25	1074586.029	1604	1553	1611
1606	1	31	1021242.442	1605	1600	1609
1605	1	75	972172.096	1602	1596	1606
1604	11	37	931639.808	1597	1591	1607
1603	44	241	891695.275	1592	1580	1610
1602	1	45	851818.731	1512	1598	1605
1601	2	58	819760.959	1590	1583	1609
1600	31	38	789447.553	1536	1593	1606
1599	84	105	761427.635	1567	1586	1613
1598	45	211	736227.006	1589	1574	1602

Figure 37. Distance between fusion coefficients in relation to number of clusters.

As it can be seen in the line graph, there is a clear, sharp elbow at cluster 4, after which the distance between the fusion coefficients starts to flatten out, thereby indicating little information is gained by adding more clusters. Taken together, both the dendrogram and the plotted distance indicate a possible 4-cluster solution as the optimal one. However, this is only an indication. Therefore, it is necessary to verify how much information is gained from five clusters and lost from three clusters (Staples & Biber, 2015). In order to accomplish this, a new cluster analysis was run for the 3-, 4-, and 5-cluster solutions.

2.5.3. Hierarchical Cluster Analysis (HCA) – Multiple clusters

A new cluster analysis, grouping the data into three different cluster solutions, was run following the same steps described above, that is, by using the path *Analyze* > *Classify* > *Hierarchical Cluster Analysis*. All variables and choices for statistics, methods, plots remained the same. The exception was for the *Save* dialogue box (Figure 38), where the range of solutions was indicated under *Cluster Membership*. In this case, the minimum number of clusters was indicated as 3 and the maximum, as 5.

Figure 38. SPSS v.23 – Hierarchical Cluster Analysis – Save cluster membership.

The results were the same as previous the hierarchical cluster analysis, but now three columns were added to the dataset, indicating the cluster membership for each node according to each solution. In order to further analyze each cluster solution and to inform the decision on the optimal number of clusters, the mean factor scores

were computed by running a one-way ANOVA for each cluster solution. This was accomplished by following the path *Analyze* > *Compare Means* > *One-way ANOVA*. The 11 factors were added to the Dependent List and, one at a time, the cluster solution was added as the Factor (Figure 39). *Descriptives* was checked under the *Options* dialogue box so that the means and standard deviations for the clusters in each factor could be obtained.

Figure 39. SPSS v.23 – Comparing means – One-way ANOVA.

The ANOVA showed significant mean differences for all 11 factors in the three cluster solutions that were extracted. The composition of two clusters remained the same in all three solutions: Clusters 2 and 3 in the 3-cluster solution; Clusters 2 and 4 in the 4-cluster solution; and Clusters 3 and 5 in the 5-cluster solution. Cluster 1 in the 3-cluster solution was divided into Clusters 1 and 3 in the 4-cluster solution. Cluster 1 in the 4-cluster solution was divided into Clusters 1 and 2 in the 5-cluster solution, further differentiating the nodes. Each solution was interpreted based on the nodes that comprised each cluster and the mean factor scores for each cluster. The 5-cluster solution was deemed the most interpretable and informative (See descriptive statistics, ANOVA results, and means plots for the 5-cluster solution in Appendix 5).

In addition to the HCA reported above, an HCA was run for each of the 11 factors to support interpretation of the nodes with the highest factor scores in each dimension. Thus, the same cluster analysis procedure described above was also conducted for the 11 factors identified in the exploratory factor analysis (Dendrograms and line graphs for the 11 HCAs can be found in Appendix 6.). The ANOVAs for all 11

HCAs showed significant mean differences in each of the cluster solutions tested. The decision on the final number of clusters for each of the 11 HCAs was based on the nodes that comprised each cluster and the mean factor scores for the cluster.

For Factor 1, the dendrogram showed a possible 2- or 3-cluster solution and the line graph showed a clear, sharp elbow at 2 clusters, so the cluster analysis was carried out for 2 and 3 clusters. Cluster 1 remained the same in both solutions and cluster 2 in the 2-cluster solution was divided into clusters 2 and 3 in the 3-cluster solution. The most interpretable solution seemed to be the 2-cluster solution.

The dendrogram for Factor 2 also showed a possible 2- or 3-cluster solution and the line graph showed a clear, sharp elbow at 2 clusters as well as a smaller, less sharp elbow at 4 clusters hence a multiple clusters analysis was carried out for 2, 3, and 4 clusters. Cluster 1 remained consistent in the 2- and 3-cluster solutions and it was divided into 2 clusters in the 4-cluster solution. Cluster 2 in the 2-cluster solution was divided into clusters 2 and 3 in the 3-cluster solution and clusters 3 and 4 in the 4-cluster solution. After interpretation, the 4-cluster solution was deemed the final one.

For Factor 3, the dendrogram showed a possible 3-cluster solution and the line graph, a clear sharp elbow at 3 clusters. Multiple cluster analysis was carried out with 2, 3, and 4 clusters to verify how much information was lost or gained by having fewer or more clusters, respectively. Cluster 1 remained the same in all 3 solutions. Cluster 2 in the 2-cluster solution was divided into clusters 2 and 3 in the 3-cluster solution. Cluster 2 in the 3-cluster solution was the same as cluster 2 in the 4-cluster solution and cluster 3 in the 3-cluster solution was divided into clusters 3 and 4 in the 4-cluster solution, with only 2 nodes in cluster 4. Therefore, the 3-cluster solution was considered the most appropriate.

According to the dendrogram for Factor 4, there was a possible 2-cluster solution. The line graph showed a clear, sharp elbow at 2 clusters and a softer elbow at 4 clusters. A cluster analysis was done for 2, 3, and 4 cluster solutions. Cluster 1 in the 2- and 3-cluster solutions remained the same, and it was divided into clusters 1 and 2 in the 4-cluster solution. Cluster 2 in the 2-cluster solution was divided into clusters 2 and 3 in the 3-cluster solution and clusters 3 and 4 in the 4-cluster solution. Thus, for Factor 4, the most interpretable solution was comprised of 2 clusters.

For Factor 5, the dendrogram showed a possible 2- or 3-cluster solution and the line graph showed a soft elbow at 3 clusters, so the cluster analysis was carried out for 2 and 3 clusters. Cluster 1 in the 2-cluster solution was divided into clusters 1

and 2 in the 3-cluster solution and cluster 2 in the 2-cluster solution was the same as cluster 3 in the 3-cluster solution. The most interpretable solution seemed to be the 3-cluster solution.

The dendrogram for Factor 6 showed a possible 2-cluster solution and the line graph showed a clear elbow at 2 clusters, so the cluster analysis was carried out for 2 and 3 clusters. Cluster 1 in the 2-cluster solution was divided into clusters 1 and 2 in the 3-cluster solution and cluster 2 in the 2-cluster solution was the same as cluster 3 in the 3-cluster solution. The 3-cluster solution was considered the most interpretable and informative.

In Factor 7, the dendrogram showed a possible 2- or 4-cluster solution and, in the line graph, there was a clear, sharp elbow at 2 clusters and a softer elbow at 4 clusters. Cluster 1 in the 2-cluster solution was divided into clusters 1 and 3 in the 3- and 4-cluster solutions. Cluster 2 in the 2-cluster solution was the same as cluster 2 in the 3-cluster solution, but it was divided into clusters 2 and 4 in the 4-cluster solution. The 3-cluster solution was considered the most appropriate.

The dendrogram for Factor 8 showed a possible 2- or 3-cluster solution and the line graph showed soft elbows at 2 and 3 clusters; therefore, a multiple clusters analysis was carried out for 2, 3, and 4 clusters. Similar to Factor 7, cluster 1 in the 2-cluster solution was divided into clusters 1 and 3 in the 3- and 4-cluster solutions. Cluster 2 in the 2-cluster solution was the same as cluster 2 in the 3-cluster solution, but it was divided into clusters 2 and 4 in the 4-cluster solution. The 3-cluster solution also seemed to be the most appropriate.

For Factor 9, the dendrogram showed a possible 2- or 3-cluster solution and the line graph showed a clear elbow at 3 clusters, so the cluster analysis was carried out for 2, 3 and 4 clusters. Cluster 1 in the 2-cluster solution was divided into clusters 1 and 2 in the 3- and 4-cluster solutions. Cluster 2 in the 2-cluster solution was the same as cluster 3 in the 3-cluster solution, but it was divided into clusters 3 and 4 in the 4-cluster solution. The most interpretable solution was considered to be the 3-cluster solution.

There was a possible 2-cluster solution according to the dendrogram and line graph for Factor 10, so the cluster analysis was carried out for 2 and 3 clusters. Cluster 1 in the 2-cluster solution was divided into clusters 1 and 2 in the 3-cluster solution and cluster 2 in the 2-cluster solution was the same as cluster 3 in the 3-cluster solution. The 3-cluster solution seemed to be the most interpretable.

Finally, for Factor 11, the dendrogram showed a possible 3-cluster solution and there was a clear, sharp elbow at 3 clusters in the line graph. Therefore, a multiple clusters analysis was carried out for 2, 3, and 4 clusters. Cluster 1 remained the same in all three solutions. Cluster 2 in the 2-cluster solution was divided into clusters 2 and 3 in the 3-cluster solution and clusters 2, 3 and 4 in the 4-cluster solution. The most interpretable solution seemed to be the 3-cluster solution.

Table 4 summarizes the final cluster solutions for the 12 HCAs carried out.

Table 4. Final cluster solutions for each of the HCAs.

Factor	Final Cluster Solution
All factors	5 clusters
Factor 1	2 clusters
Factor 2	4 clusters
Factor 3	3 clusters
Factor 4	2 clusters
Factor 5	3 clusters
Factor 6	3 clusters
Factor 7	3 clusters
Factor 8	3 clusters
Factor 9	3 clusters
Factor 10	3 clusters
Factor 11	3 clusters

For the overall HCA as well as for each of the factors, the clusters were interpreted in an attempt to identify possible node typologies that comprise each factor, further characterizing the dimensional collocational networks that emerged from the factor analysis. The interpretation was based on semantic grouping of nodes, analysis of collocations that loaded on the factors and clusters, and analysis of text excerpts. In addition, WordNet definitions, listed with the support of a script in Python, were used to support the finding of node categories. In addition to the nodes that comprised each cluster, descriptive statistics were also used to further support interpretation of the collocational dimensions identified in the factor analysis. For each factor, the frequencies per text of the 30 nodes with the highest positive factor scores and of the 30 nodes with the highest negative factor scores were computed as an equivalent to a “dimension score” for each text in an effort to understand the nature of the dimensions with respect to the texts and discipline subfields that are present in the corpus.

This chapter has described in detail the methodological steps carried out in this study, ranging from the compilation and preparation of corpus texts and extraction of collocations to the multivariate statistical procedures used, namely exploratory factor analysis and hierarchical cluster analysis. In the next chapter, the results are detailed, and the dimensional collocational networks and node types identified in each of the factors are described and interpreted.

Chapter 3. Results

The results of the exploratory factor analysis and hierarchical cluster analysis detailed in Chapter 2 are presented in this chapter. For each collocational dimension identified, the results of the exploratory factor analysis are presented and discussed, followed by the results and discussion of the cluster analysis.

3.1. Collocational Dimensions of Written Academic English

As described in the previous chapter, an exploratory factor analysis, with principal axis as the extraction method, was conducted in a dataset containing 1,615 nodes as observations, 323 collocates as variables, and logDice statistic as cell values. The nodes and collocates were extracted from the Social, Behavioral, and Economic Sciences subcorpus of the CAT, comprising 230 texts (7,869,573 tokens; 709,194 types). An 11-factor solution was extracted, giving rise to the following collocational dimensions:

1. Finance and Economics
2. Classification
3. Social and Cultural Development
4. Exerting Influence and Shaping Outcomes
5. Empirical Research Methods
6. Local and Global Society
7. Reporting Research
8. Individuals in Society
9. Gaining Knowledge and Developing Concepts
10. Social Welfare
11. Marking Time and Sectioning

The fact that meaningful and interpretable collocational dimensions were extracted can be seen as evidence of the existence of extended collocational networks, with sets of collocates showing a tendency to co-occur with given sets of node words. The 11 collocational dimensions extracted are described in detail in the following subsections.

3.1.1. Collocational Dimension 1: Finance and Economics

In total, 42 collocates loaded on Dimension 1, entitled Finance and Economics, comprising 1,771 unique collocations[7] across 527 unique nodes. The 1,771 unique collocations represent the extended collocational network for the 42 collocates co-occurring across 527 nodes. That is, the interaction of 42 words on the one hand (the collocates) with 527 words on the other (the nodes) is giving rise to 1,771 combinations. As it can be seen on Table 5, there is a predominance of collocates as nouns (69%), with verbs and adjectives corresponding to 17% and 14%, respectively.

As the name suggests, the major semantic fields of the collocates that loaded on this dimension are finance and economics. The nodes with the highest positive factor scores are also from the same domains, as shown in the sample in Table 6. All 44 collocates (100%) can be found in the top 100 nodes with the highest positive factor scores, with more than two-thirds (28 of the 44; 64%) of them among the 30 highest-ranking nodes. The sample of nodes with the highest negative factor scores shown in the same table are related to markers of time and text structure. However, it should be highlighted that the majority of the nodes (1,258 nodes; 78%) had a negative factor score, and a large proportion of them had the same negative score (1,088 nodes).

Table 5. Composition of Dimension 1.

Dimension 1		
Collocate	**Part of Speech**	**Loading[8]**
wage	Noun	0.57
low	Adjective	0.56
price	Noun	0.56
rate	Noun	0.55
increase	Verb	0.55
average	Adjective	0.53
high	Adjective	0.52
increase	Noun	0.52
income	Noun	0.49
investment	Noun	0.49
capital	Noun	0.48
cost	Noun	0.48
level	Noun	0.48

[7] The unique collocations from each dimension can be found in Appendix 8

[8] Loadings in every table have been rounded to two decimal spaces.

return	Noun	0.47
percent	Noun	0.47
reduce	Verb	0.46
profit	Noun	0.46
share	Noun	0.45
expect	Verb	0.45
firm	Noun	0.44
total	Adjective	0.44
risk	Noun	0.44
tax	Noun	0.42
production	Noun	0.42
demand	Noun	0.42
growth	Noun	0.40
pay	Verb	0.39
market	Noun	0.38
interest	Noun	0.37
labor	Noun	0.36
associate	Verb	0.34
less	Adjective	0.33
amount	Noun	0.33
worker	Noun	0.32
value	Noun	0.32
relative	Adjective	0.31
raise	Verb	0.31
lead	Verb	0.30
size	Noun	0.28
asset	Noun	0.25
number	Noun	0.24
money	Noun	0.22

Table 6. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 1.

Dimension 1

Node	Part of Speech	Factor Scores
high	Adjective	191.50
increase	Verb	189.52
increase	Noun	173.96
price	Noun	172.20
rate	Noun	164.27
low	Adjective	147.95
firm	Noun	141.45
cost	Noun	132.10
reduce	Verb	126.45
total	Adjective	108.52
wage	Noun	106.31
market	Noun	106.20
income	Noun	104.68
level	Noun	103.12
average	Adjective	101.12
percent	Noun	98.44

value	Noun	89.27
pay	Verb	88.43
labor	Noun	88.07
capital	Noun	87.78
risk	Noun	86.21
tax	Noun	83.01
return	Noun	79.54
earn	Verb	76.94
investment	Noun	74.89
worker	Noun	74.75
large	Adjective	71.08
production	Noun	68.58
amount	Noun	67.64
raise	Verb	64.30
next	Adjective	-6.73
last	Adjective	-6.73
month	Noun	-6.73
week	Noun	-6.73
early	Adjective	-6.73
first	Adjective	-6.73
decade	Noun	-6.73
begin	Verb	-6.73
previous	Adjective	-6.73
end	Noun	-6.73
chapter	Noun	-6.73
section	Noun	-6.73
detail	Noun	-6.73
past	Adjective	-6.73
late	Adjective	-6.73
century	Noun	-6.73
examine	Verb	-6.73
discuss	Verb	-6.73
describe	Verb	-6.73
see	Verb	-6.73
beginning	Noun	-6.73
several	Adjective	-6.73
twentieth	Adjective	-6.73
follow	Verb	-6.73
let	Verb	-6.73
second	Adjective	-6.73
half	Noun	-6.73
step	Noun	-6.73
short	Adjective	-6.73
night	Noun	-6.73

An analysis of the collocations made up of nodes and collocates with the highest loadings and factor scores and the context in which they appear in the corpus reinforced the perception that these collocations are supporting discussion about topics related to the large domains of finance and economics, including, but not limited to: (1) job market (e.g., *wage, income, firm, labor, worker*), (2) investments (e.g., *investment*, *capital, risk, return, profit, share, interest, asset, value*), and (3) market economy and business (e.g., *price, cost, production, demand, growth, pay, market, money, tax*). The following excerpts exemplify how the collocations are used to build aboutness and subject-matter in these domains.

*[1] A **firm** may decide to **pay** a **wage** above the equilibrium level to maintain a healthy workforce. […] A second efficiency-wage theory, which is more relevant for developed countries, holds that <u>high wages</u> **reduce labor** turnover. […] The more a **firm pays** its **workers**, the greater is their incentive to stay with the firm. By **paying** a **high wage**, a **firm reduces** the frequency at which its workers quit, thereby decreasing the time and money spent **hiring** and training new **workers**. (5_7_18.txt)*

*[2] **Capital gains** and **losses**, on the other hand, are treated differently. Assets such as stocks, bonds, and real estate are defined as **capital assets**. When you buy a **capital asset** and later sell it for more than you paid, you **earn** a **profit** that is called a **capital gain**; when you suffer a loss, it is called a **capital loss**. If you **held** the **asset** for a year or less, you will have a short-term gain or loss, while if you held it for more than a year, you will have a long-term gain or loss. (5_7_21.txt)*

*[3] How much will each **firm produce**? Each firm cares about the **market price** when it selects its **production level**. Because **market price** depends on the **total production** of both **firms**, the amount that Hynix desires to produce depends on how much it expects Micron to produce (and vice versa). Cournot investigated production under a simple set of expectations. Each firm "guesses" how much the other **firm** will **produce** and believes that its rival will stick to this **level** of **output**.14 Each firm's optimal **level** of **production** is the best response to the **level** it **expects** its rival to choose. (5_7_24.txt)*

Besides identifying extended collocational networks for nodes and collocates, it was also speculated that collocational patterns may fulfill functional roles in structuring discourse. An example of this may be the fact that all the adjective collocates and the highest scoring adjective nodes—such as *low, average, high, total, less, relative, large*— offer an order of magnitude, extent and size to economical concepts, as shown in the examples that follow. The attributive adjectives here are modifying abstract/process or quantity nouns.

*[4] In any given year, the **average number** of dirty patents per firm is 0.23 and the **average number** of clean patents is 0.08. (5_6_136.txt)*

*[5] For example, drug addicts are willing to pay a **high price** for heroin. Yet we would not say that addicts get a large benefit from being able to buy heroin at a **low price** (even though addicts might say they do). (5_7_19.txt)*

*[6] Unfortunately, this policy works only so long as people keep their jobs — and **higher** wage **rates** under given conditions, especially deflation, mean **lower employment**. Therefore **higher** real wage **rates** per hour did not translate into **higher** aggregate **earnings** for labor, and so provided no basis for the **higher** aggregate **demand** that both presidents expected. (5_7_20.txt)*

Moreover, the notion of size, quantity and magnitude is also reflected in the collocations with *increase* (both as a verb and a noun), *rate, reduce, level, size, number, amount, raise,* and *percent*, as the following excerpts exemplify.

*[7] For a woman, the reward for attending college rose from a 35 percent **increase** in **earnings** in 1975 to an 81 **percent increase** in 2011. (5_7_19.txt)*

*[8] a budget deficit **raises** interest **rates** and crowds out investment; the resulting reduction in the capital stock is part of the burden of the national debt on future generations. (5_7_18.txt)*

*[9] Another benefit of production-related value chain commonalities is the ability to consolidate production into a smaller number of plants and significantly **reduce** overall production **costs**. (5_7_29.txt)*

As mentioned in Chapter 2, a hierarchical cluster analysis was conducted in order to categorize nodes and support interpretation of the collocational dimensions. For Dimension 1, the cluster analysis returned a 2-cluster solution, with 1,537 nodes

on Cluster 1 and 78 nodes on Cluster 2. Only 449 nodes in Cluster 1 formed collocations with collocates that loaded in this dimension. The nodes in Cluster 2 corresponded exactly to the 78 nodes with the highest factor scores in Dimension 1.

The main categories that emerge for nodes in Cluster 2 can be defined as: (1) investment (e.g., *investment, return, capital, profit, stock, interest, value, risk, share, money, financial, asset*); (2) quantification (e.g., *high, low, increase, reduce, raise, average, level, amount, number, size, total, rate, grow, relative*); (3) labor (e.g., *wage, income, earn, market, labor, worker*); and (4) market (e.g., *pay, supply, demand, production, product, sale, productivity, firm, cost*). Although the 78 nodes from Cluster 2 seem to be the ones driving the characterization of this dimension, there are other 449 nodes co-occurring with the collocates that loaded in this dimension, which reflect similar themes to a certain extent, but seem to characterize more possible outcomes or factors—mostly adverse—that might affect economics and finance, such as, for example: (1) negative factors and outcomes (e.g., *inequality, unemployment, poverty, inflation, decline, uncertainty, mortality, crime, punishment, conflict, poor, volatility, incarceration, divorce*); and (2) neutral and positive factors and outcomes (e.g., *college, education, culture, diversity, behavior, traditional, idea, relationship, influence, adjustment, incentive, construction, potential, productive*).

3.1.2. Collocational Dimension 2: Classification

Dimension 2, Classification, consists of 25 collocates, all of them nouns, as shown in Table 7. A total of 824 unique collocations, across the 25 collocates and 401 unique nodes, are present in this dimension. The 824 unique collocations represent the extended collocational network for the 25 collocates co-occurring across 401 nodes.

In this dimension, lexical and functional structuring of discourse becomes even more evident. According to Biber et al., “classification is an important aspect of academic procedure and discourse” (1999, p. 256). Most of the collocates in this dimension are species nouns, which tend to occur in higher frequencies in academic prose. Another finding that corroborates this is that the nodes with the highest positive factor scores in this dimension are mostly relational classifiers functioning as attributive adjectives (Table 8). Only 6 of the 25 collocates (24%) can be found in the top 100 nodes with the highest positive factor scores, with none of them being among

the 30 highest-ranking nodes. Most of the nodes with the highest negative factor scores shown in Table 8 are related to markers of time and text structure or sectioning. Similar to Dimension 1, 75% of the nodes (1,214 nodes) had a negative factor score; however, all of them had the exact same value (-3.57).

Table 7. Composition of Dimension 2.

Dimension 2		
Collocate	**Part of Speech**	**Loading**
kind	Noun	0.69
feature	Noun	0.57
aspect	Noun	0.52
type	Noun	0.52
area	Noun	0.51
reason	Noun	0.47
characteristic	Noun	0.44
way	Noun	0.42
thing	Noun	0.42
place	Noun	0.41
form	Noun	0.40
product	Noun	0.34
condition	Noun	0.33
time	Noun	0.32
source	Noun	0.32
factor	Noun	0.31
activity	Noun	0.29
category	Noun	0.28
set	Noun	0.28
problem	Noun	0.27
question	Noun	0.27
word	Noun	0.26
task	Noun	0.25
company	Noun	0.22
function	Noun	0.20

Table 8. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 2.

Dimension 2		
Node	**Part of Speech**	**Factor Scores**
other	Adjective	141.17
different	Adjective	138.34
such	Adjective	112.21
many	Adjective	108.13
particular	Adjective	102.24

specific	Adjective	92.54
important	Adjective	76.16
various	Adjective	72.88
certain	Adjective	72.58
same	Adjective	66.42
several	Adjective	55.96
include	Verb	55.18
new	Adjective	53.65
good	Adjective	50.63
key	Adjective	42.33
major	Adjective	39.42
social	Adjective	38.19
variety	Noun	37.67
simple	Adjective	35.67
use	Verb	34.88
human	Adjective	34.30
define	Verb	33.73
associate	Verb	32.50
value	Noun	31.49
people	Noun	30.01
first	Adjective	29.61
create	Verb	29.49
similar	Adjective	28.49
learn	Verb	28.44
response	Noun	27.17
next	Adjective	-3.57
last	Adjective	-3.57
month	Noun	-3.57
week	Noun	-3.57
decade	Noun	-3.57
previous	Noun	-3.57
end	Noun	-3.57
chapter	Noun	-3.57
section	Noun	-3.57
detail	Adjective	-3.57
past	Noun	-3.57
late	Noun	-3.57
century	Adjective	-3.57
beginning	Verb	-3.57
twentieth	Noun	-3.57
follow	Noun	-3.57
let	Noun	-3.57
half	Noun	-3.57
step	Noun	-3.57
night	Noun	-3.57

history	Adjective	-3.57
method	Noun	-3.57
entire	Noun	-3.57
billion	Verb	-3.57
stage	Noun	-3.57
introduce	Noun	-3.57
discussion	Noun	-3.57
book	Noun	-3.57
contain	Noun	-3.57
nineteenth	Noun	-3.57

Thus, the collocates that loaded on Dimension 2 can be said to be referring to classification or attributes of entities. The nodes that collocate with highest strength are classifier attributive adjectives, as evident in the examples that follow this paragraph. The collocates in this dimension can be divided into: species nouns (*kind, feature, aspect, type, way, form, set, category*), cognitive nouns (*reason, problem, question, function*), group/institution nouns (*product, company*), abstract/process nouns (*source, factor, characteristic, condition, activity, task*), and technical/concrete and quantity nouns (*area, thing, place, word, time*).

[10] Ethnologists often focus on a particular ***aspect*** *of* ***culture****, such as religious or economic practices. (5_7_7.txt)*

[11] ***Several features*** *distinguish rituals from other* ***kinds*** *of* ***behavior*** *(Rappaport 1974). Rituals are formal — stylized, repetitive, and stereotyped. (5_7_6.txt)*

[12] research suggests that people engage in discretionary behaviors at ***different times*** *for* ***different reasons*** *(e.g., Dalal et al., 2009). (5_6_192.txt)*

[13] Because a demand-side shock involves heterogeneous changes across different ***product markets*** *(demand for* ***certain products*** *increases, […]) (5_6_35.txt)*

The cluster analysis revealed 4 clusters, containing 1,214, 344, 51, and 6 nodes, respectively. The 401 nodes that make up collocations found in this dimension loaded only on clusters 2, 3, and 4. None of the nodes found in Cluster 1 form collocations (given the parameters used in this study) with the with the collocates that loaded in Dimension 2.

The nodes in Cluster 4—*different, other, many, such, particular,* and *specific*—are the nodes with the highest factor scores. They appear to function as categorization of elements. Nodes in Cluster 3 may be grouped into the following categories: (1) relational classification (e.g., *various, certain, several, first, second, same*); (2) significance (e.g., *important, major, key, basic, simple*); (3) life (e.g., *human, people, life, live, social, identify, behavior, response*); (4) productivity (e.g., *demand, production, economic, work, information*); and (5) actions (e.g., *use, take, create, make, perform, use*). The major categories in Cluster 2 seem to be the following: (1) analysis (e.g., *consider, data, analysis, research, present, discuss, determine, measure, describe, evidence, parameter*); (2) society aspects (e.g., *environmental, language, political, culture, cultural, biological, demographic, medical, racial, sexual, gender, marriage, society*); (3) economic activity (e.g., *buy, sell, provide, product, market, price, money, develop, utility*); (4) relevance (e.g., *distinctive, complex, obvious, right, distinct, rich, relevant, critical, great*); (5) geography (e.g., *land, local, world, region, area, geographic, agricultural, rural, urban*); and (6) expression (e.g., *say, speak, address, call, ask, represent, meaning, know, word, cognitive*).

In this dimension, the collocational patterns identified seem to structure discourse more functionally than lexically. Topics, aboutness, and subject matter become secondary structures, with lexicogrammatical characteristics of academic prose being highlighted.

3.1.3. Collocational Dimension 3: Social and Cultural Development

Dimension 3, Social and Cultural Development, is comprised of 31 collocates, as shown in Table 9. Overall, 1,171 unique collocations were formed across 474 unique nodes and the collocates that characterize this dimension. The 1,171 unique collocations represent the extended collocational network for the 31 collocates co-occurring across 474 nodes. Only 6% of the collocates are verbs, 16% are adjectives, and the remaining 77% are nouns.

Table 10 shows the nodes with the highest positive and negative factor scores in this dimension. Nodes with high positive factor scores reflect the same semantic domains as the collocates. Only 4 of the 31 collocates are not found in the top 100 nodes with the highest positive factor scores, that is, there are 87% of collocates (27

out of 31) in common with the highest-ranking nodes. Of those, 52% (14 of the 27) are in the top 30 nodes. There were 1,177 nodes (73%) with negative factor scores and 1,141 of those had the same score (-4.68).

Table 9. Composition of Dimension 3.

Dimension 3		
Collocate	**Part of Speech**	**Loading**
theory	Noun	0.44
social	Adjective	0.41
process	Noun	0.38
culture	Noun	0.36
behavior	Noun	0.35
environment	Noun	0.34
cultural	Adjective	0.34
development	Noun	0.33
system	Noun	0.33
context	Noun	0.31
knowledge	Noun	0.30
language	Noun	0.30
experience	Noun	0.29
idea	Noun	0.29
skill	Noun	0.29
structure	Noun	0.28
basic	Adjective	0.27
status	Noun	0.27
complex	Adjective	0.27
race	Noun	0.27
approach	Noun	0.26
pattern	Noun	0.25
human	Adjective	0.25
role	Noun	0.24
force	Noun	0.24
practice	Noun	0.24
ability	Noun	0.23
maintain	Verb	0.23
belief	Noun	0.22
help	Verb	0.21
power	Noun	0.21

Table 10. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 3.

Dimension 3		
Node	**Part of Speech**	**Factor Scores**
social	Adjective	154.23
cultural	Adjective	107.48
understand	Verb	80.71
human	Adjective	79.64
different	Adjective	75.84
learn	Verb	72.44
economic	Adjective	71.42
create	Verb	69.82
political	Adjective	66.17
base	Verb	62.38
system	Noun	62.20
behavior	Noun	61.78
shape	Verb	57.36
change	Noun	54.67
develop	Verb	54.32
affect	Verb	54.16
language	Noun	52.48
knowledge	Noun	48.88
new	Adjective	48.78
culture	Noun	47.88
basic	Adjective	47.00
theory	Noun	45.36
work	Noun	42.57
complex	Adjective	42.04
power	Noun	41.85
own	Adjective	41.05
aspect	Noun	40.43
development	Noun	39.29
such	Adjective	39.26
idea	Noun	38.87
next	Adjective	-4.68
last	Adjective	-4.68
month	Noun	-4.68
week	Noun	-4.68
decade	Noun	-4.68
end	Noun	-4.68
chapter	Noun	-4.68
section	Noun	-4.68
detail	Noun	-4.68
late	Adjective	-4.68
century	Noun	-4.68
beginning	Noun	-4.68

twentieth	Adjective	-4.68
let	Verb	-4.68
half	Noun	-4.68
night	Noun	-4.68
billion	Noun	-4.68
discussion	Noun	-4.68
book	Noun	-4.68
contain	Verb	-4.68
nineteenth	Adjective	-4.68
run	Noun	-4.68
generation	Noun	-4.68
end	Verb	-4.68
distance	Noun	-4.68
interval	Noun	-4.68
middle	Noun	-4.68
career	Noun	-4.68
implication	Noun	-4.68
finding	Noun	-4.68

The collocations found in this dimension reflect themes mainly related to: social and cultural aspects of life in society (e.g., *social behavior, social context, social status, cultural practice, cultural force, cultural process, different culture, economic system*); human development (e.g., *develop ideas, develop system, behavior change, environment change, shape experience, cultural adaptation*), cognition (e.g., *ability* to *learn*, *learn skills, understand process, cognitive ability, cognitive process*), and literacy (e.g., *own language, learn language, share knowledge, language development*). These topics are exemplified in the following excerpts.

*[14] They cannot achieve the extremely complex **human language** system that is perhaps the most distinct aspect of **human culture** and that enables us to store and pass on huge quantities of information not embedded in our genes. (5_7_1.txt)*

*[15] **culture** has two interacting **roles**: as the means by which we respond adaptively to our **natural environments** and as the means by which we respond to our **cultural environments** — that is, the ways in which we maintain the integrity and meaning of our **cultural systems** (see Chapter 4 and Figure 4.4). A **cultural behavior**, then, has both practical value and meaning. Now we*

must try to put all these ideas together and see how they work when we look not just at individual aspects of a ***cultural system*** *but at the* ***cultural system*** *as a whole. (5_7_3.txt)*

[16] Beginning in infancy, children depend on others to ***learn language****, to learn how to act in culturally appropriate ways, and to engage with the physical environment effectively. (5_6_11.txt)*

[17] Modern advances in transportation and communication can break through the isolation of many peoples, just as other technological advances can mitigate, or sometimes even eliminate, the handicaps of various other kinds of geographic impediments to economic and ***social development****. But what these historically recent advances cannot do is retroactively erase the effects of thousands of years of different* ***cultural development*** *that took place where there were serious geographic restrictions […]. (5_7_23.txt)*

A 3-cluster solution was extracted for this dimension. Cluster 3 contained the 40 nodes with the highest factor scores in this dimension, thus they can be grouped into categories that reflect similar topics as those identified above, that is, development and learning (e.g., *learn, develop, create, understand, cognitive, idea, skill, research, development, theory, share, behavior, language, knowledge*); and cultural, political and socioeconomic aspects and systems (e.g., *culture, power, process, structure, society, belief, system, shape, base, social, cultural, economic, political, human, change, work*). There were 173 nodes in Cluster 2, which seem to reflect aspects required to facilitate development in society, that is, the main categories reflect: (1) participants in social and cultural development (e.g., *people, child, woman, individual, group, class, institution, interaction, role, race, status, relationship, gender, institution*); (2) change and adaptation (e.g., *position, change, maintain, traditional, evolution, vary, evolutionary, variation, growth, adapt, adaptation, past, history, recent*); (3) awareness and thinking (e.g., *reflect, explain, study, describe, information, concept, pattern, examine, define, identify, observe*); (4) taking action (e.g., *produce, provide, require, acquire, gain, make, achieve, take, give, contribute, work*); and (5) influence (e.g., *apply, determine, influence, engage, involve, include, suggest, emphasize, control, allow, attitude*).

The remaining 1,402 nodes loaded on Cluster 1; however, out of those, only 261 nodes formed collocations with the collocates present in this dimension. The

nodes in this dimension seem to follow the same broad semantic categories as the other clusters, providing further lexical differentiation.

3.1.4. Collocational Dimension 4: Exerting Influence and Shaping Outcomes

In total, 34 collocates loaded on Dimension 4, entitled Exerting Influence and Shaping Outcomes, comprising 1,227 unique collocations across 457 unique nodes. The 1,227 unique collocations represent the extended collocational network for the 34 collocates co-occurring across 457 nodes. As it can be seen on Table 11, there is a predominance of collocates as verbs (50%) and adjectives (41%), with nouns corresponding to only 8%.

As it can be seen in the sample in Table 12, the nodes with the highest factor scores are mostly nouns, representing domains of influence and/or processes/contexts. Eleven of the 34 collocates (32%) can be found in the top 100 nodes with the highest positive factor scores. Out of those, only 3 are found in the top 30 highest-ranking nodes. A total 1,206 nodes (75%) had negative factor scores, where 1,159 nodes had the same value (-5.0).

Table 11. Composition of Dimension 4.

Dimension 4		
Collocate	**Part of Speech**	**Loading**
influence	Verb	0.48
affect	Verb	0.48
particular	Adjective	0.46
change	Verb	0.43
different	Adjective	0.42
shape	Verb	0.41
certain	Adjective	0.40
such	Adjective	0.36
change	Noun	0.36
response	Noun	0.33
occur	Verb	0.30
various	Adjective	0.30
specific	Adjective	0.30
reflect	Verb	0.29
depend	Verb	0.28
determine	Verb	0.28

political	Adjective	0.28
create	Verb	0.28
individual	Adjective	0.27
follow	Verb	0.26
economic	Adjective	0.26
understand	Verb	0.26
outcome	Noun	0.26
key	Adjective	0.26
important	Adjective	0.26
see	Verb	0.26
improve	Verb	0.25
study	Verb	0.25
measure	Verb	0.25
major	Adjective	0.24
cause	Verb	0.24
vary	Verb	0.23
same	Adjective	0.23
simple	Adjective	0.20

Table 12. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 4.

Dimension 4		
Node	**Part of Speech**	**Factor Scores**
behavior	Noun	134.03
way	Noun	111.81
factor	Noun	100.25
change	Noun	90.51
variable	Noun	88.70
culture	Noun	83.53
policy	Noun	73.28
process	Noun	71.28
social	Adjective	70.35
condition	Noun	65.29
group	Noun	63.10
structure	Noun	62.39
difference	Noun	60.18
effect	Noun	60.14
characteristic	Noun	59.98
people	Noun	58.69
economic	Adjective	58.57
activity	Noun	57.98
environment	Noun	57.89
cultural	Adjective	57.39
time	Noun	56.54
event	Noun	55.89

level	Noun	55.37
role	Noun	54.70
situation	Noun	54.58
price	Noun	53.80
ability	Noun	53.64
context	Noun	53.21
area	Noun	49.02
decision	Noun	48.94
last	Adjective	-5.00
week	Noun	-5.00
decade	Noun	-5.00
end	Noun	-5.00
detail	Noun	-5.00
late	Adjective	-5.00
century	Noun	-5.00
beginning	Noun	-5.00
twentieth	Adjective	-5.00
let	Verb	-5.00
half	Noun	-5.00
night	Noun	-5.00
billion	Noun	-5.00
book	Noun	-5.00
contain	Verb	-5.00
nineteenth	Adjective	-5.00
run	Noun	-5.00
generation	Noun	-5.00
end	Verb	-5.00
distance	Noun	-5.00
interval	Noun	-5.00
middle	Noun	-5.00
career	Noun	-5.00
lag	Noun	-5.00
european	Adjective	-5.00
duration	Noun	-5.00
text	Noun	-5.00
minute	Noun	-5.00
following	Adjective	-5.00
turn	Verb	-5.00

The extended collocational networks in this dimension appear to show collocates as representing actions over concepts or processes, as shown in excerpts 18-21 that follow. The main topics identified after an analysis of how the collocations are used in the corpus are: (1) creating and shaping aspects of society by exerting influence over processes and people, such as, for example, economic policies or

social and cultural behaviors; and (2) determining outcomes of situations or processes through change.

*[18] the emergence of Obama on the political scene also held the potential to **affect** the **ways** in which blacks perceived a decrease in discrimination and, therefore, an increase in the number of opportunities available to them. (5_6_177.txt)*

*[19] Financing constraints **influence** the investment **behavior** of firms just as borrowing constraints **influence** the consumption **behavior** of households. (5_7_18.txt)*

*[20] This need is especially great in developing countries where high temperatures can **cause** dramatic **changes** in life expectancy. (5_6_132.txt)*

*[21] In other words, one must be willing to question the structural arrangements that **shape** social **behavior**. When we have this imagination, we begin to see the solutions to social problems not in terms of **changing** problem **people** but in **changing** the **structure** of society. (5_7_16.txt)*

The cluster analysis for Dimension 4 revealed a 2-cluster solution. Cluster 1 contained 1,512 nodes and Cluster 2, 103 nodes. As expected, the nodes in Cluster 2 correspond to the nodes with the highest factor scores in this dimension. It may be said that the vast majority of the nodes in Cluster 2 reflect concepts or aspects that are being either modified by the action verb collocates or classified by relational adjectives as collocates. The nodes can be grouped into the following broad categories: (1) courses of action (e.g., *policy, process, activity, event, model, system, method, action, choice, question, decision, role*); (2) feature or reasoning (e.g., *factor, condition, characteristic, situation, context, aspect, pattern, circumstance, feature, approach, issue, question, reason, perception*); (3) society (e.g., *society, behavior, culture, cultural, social, economic, political, environment, language, life*); (4) members of society and place (e.g., *people, group, institution, human, population, individual, part, component, country, region, world*); (5) economic activity (e.g., *demand, market, force, income, price, product, practice, task, function*); (6) scale of intensity, amount or quality (e.g., *level, rate, size, amount, degree, measure, extent, difference, performance, impact, effect, quality, value, number*); and (7) general classification (e.g., *type, area, category, kind, such, other*).

Out of the total number of nodes in Cluster 1, only 354 formed collocations with the collocates that loaded in this dimension. In addition to further differentiation of the categories identified in Cluster 2, there were also references to time and references internal to the text (e.g., *early, last, next, chapter, section, page, first, paper*).

3.1.5. Collocational Dimension 5: Empirical Research Methods

Dimension 5, Empirical Research Methods, is comprised of 27 collocates, which are shown in Table 13. In total, there were 1,069 unique collocations formed across 450 unique nodes and the 27 collocates from this dimension. The 1,069 unique collocations represent the extended collocational network for the 27 collocates co-occurring across 450 nodes. In terms of part of speech, 81% of the collocates are nouns, 15% are adjectives, and only 4% are verbs.

Table 14 shows the nodes with the highest positive and negative factor scores in this dimension. Only 4 of the 27 collocates are not found in the top 100 nodes with the highest positive factor scores, that is, there are 85% of collocates (23 out of 27) in common with the highest-ranking nodes. Of those, 57% (13 of the 23) are in the top 30 nodes. Same as in Dimension 2, all of the nodes with negative factor scores (1,165 nodes; 72%) had the same value of -4.19.

Table 13. Composition of Dimension 5.

Dimension 5		
Collocate	**Part of Speech**	**Loading**
model	Noun	0.58
estimate	Noun	0.55
result	Noun	0.48
analysis	Noun	0.48
datum	Noun	0.47
effect	Noun	0.43
regression	Noun	0.43
equation	Noun	0.41
estimate	Verb	0.41
table	Noun	0.40
study	Noun	0.36
measure	Noun	0.35
significant	Adjective	0.32
example	Noun	0.32

experiment	Noun	0.32
variable	Noun	0.32
sample	Noun	0.30
evidence	Noun	0.27
similar	Adjective	0.26
method	Noun	0.25
positive	Adjective	0.25
test	Noun	0.25
relationship	Noun	0.24
term	Noun	0.24
participant	Noun	0.23
negative	Adjective	0.22
difference	Noun	0.21

Table 14. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 5.

Dimension 5		
Node	**Part of Speech**	**Factor Scores**
use	Verb	133.15
show	Verb	109.80
effect	Noun	87.06
provide	Verb	74.43
model	Noun	73.02
coefficient	Noun	71.79
include	Verb	70.44
indicate	Verb	68.75
base	Verb	66.47
study	Noun	63.66
present	Verb	62.75
result	Noun	61.20
regression	Noun	55.66
report	Verb	54.00
find	Verb	53.46
estimate	Verb	53.22
significant	Adjective	48.22
variable	Noun	47.84
analysis	Noun	47.27
datum	Noun	46.48
suggest	Verb	46.48
describe	Verb	46.17
give	Verb	43.80
relationship	Noun	43.69
estimate	Noun	41.78
different	Adjective	40.78
equation	Noun	39.15

first	Adjective	38.71
obtain	Verb	38.01
such	Adjective	35.52
week	Noun	-4.19
decade	Noun	-4.19
end	Noun	-4.19
detail	Noun	-4.19
late	Adjective	-4.19
century	Noun	-4.19
beginning	Noun	-4.19
twentieth	Adjective	-4.19
let	Verb	-4.19
half	Noun	-4.19
night	Noun	-4.19
billion	Noun	-4.19
book	Noun	-4.19
nineteenth	Adjective	-4.19
run	Noun	-4.19
generation	Noun	-4.19
end	Verb	-4.19
distance	Noun	-4.19
interval	Noun	-4.19
middle	Noun	-4.19
career	Noun	-4.19
lag	Noun	-4.19
european	Adjective	-4.19
duration	Noun	-4.19
text	Noun	-4.19
minute	Noun	-4.19
turn	Verb	-4.19
exception	Noun	-4.19
name	Verb	-4.19
quarter	Noun	-4.19

The collocations formed by collocates that loaded in this dimension are used to talk about empirical research methods and statistical procedures, as exemplified by the following excerpts.

*[22] We first **estimate** a linear **regression** with year fixed effects dt as follows: […] with standard errors clustered by firm. (5_6_135.txt)*

*[23] With more **independent variables** in the model, a larger **sample size** is usually needed to use the t approximation. (5_7_30.txt)*

[24] Important insights about population codes further arise from studies that simultaneously measure spiking responses and mass signals, or that perform ***comparative analysis*** *on such* ***data obtained*** *in separate experiments. (5_6_9.txt)*

[25] Our bootstrap method works as follows. A ***random sample*** *with replacement is generated from the* ***full sample*** *of countries. In this* ***random sample****, we extract the first principal component of the five Gini indicators that capture inequality in geography across ethnic lines on elevation, precipitation, temperature, distance to coast, and land quality. We then use this principal component from the* ***random sample*** *in the regression where the* ***dependent variable*** *is ethnic inequality. (5_6_140.txt)*

Although the cluster analysis for Dimension 5 revealed a 3-cluster solution, none of the nodes in Cluster 1 (1,165 nodes in total) formed collocations with the collocates that loaded in this dimension. The 450 nodes that did form collocations were identified in Cluster 2 (421 nodes) and Cluster 3 (29 nodes).

The nodes in Cluster 3 are the ones with the highest factor scores, which seem to support the collocates in reporting the results of empirical research, as shown by the group of reporting verbs found in this cluster (*show, provide, include, indicate, present, report, find, suggest, describe, give*) as well as the nodes associated with research methods and statistics (e.g., *model, coefficient, effect, study, regression, estimate, variable, analysis, datum, equation*). An example of this in the corpus is shown in the following excerpt:

[26] For the bisectable triplets, the factor range ***showed*** *a* ***significant*** *effect on the confidence reports; participants were significantly more confident about their answer if the NBTs had a small range (5_6_90.txt)*

The nodes in Cluster 2 can be grouped mainly into the following categories: (1) conduct research (e.g., *apply, test, conduct, perform, measure, take, study, examine, observe, compare, develop, generate, consider, assess, predict*); (2) report research (e.g., *see, reveal, illustrate, explain, demonstrate, yield*); (3) results (e.g., *outcome, impact, magnitude, evidence, report*); (4) methods and statistics (e.g., *correlation, parameter, linear, value, difference, method, interaction, research, error, random,*

finding, association, hypothesis, pattern); (5) study participants (e.g., *participant, participants, partner, adult, researcher, population, sample, group, student, human, individual, subject, people, child, woman*); and (6) qualification of magnitude (e.g., *good, positive, negative, strong, consistent, simple, small, large, new, classic, little, important, substantial, average, great, ordinary, full, long, short, appropriate, robust, considerable*).

3.1.6. Collocational Dimension 6: Local and Global Society

Dimension 6, Local and Global Society, consists of 29 collocates, where 62% of them are nouns, and 24% and 14% are adjectives and verbs, respectively, as shown in Table 15. A total of 1,109 unique collocations, across the 29 collocates and 427 unique nodes, are present in this dimension. The 1,109 unique collocations represent the extended collocational network for the 29 collocates co-occurring across 427 nodes.

Table 16 shows the nodes with the highest factor scores in this dimension, both positive and negative. Almost all 29 collocates (28 out of 29; 97%) can be found in the top 100 nodes with the highest positive factor scores, with 87% of them (20 of the 23) among the 30 highest-ranking nodes. In total, 1,214 nodes had negative factor scores in this dimension, with 1,188 (75%) of them having the same value of -4.44.

Table 15. Composition of Dimension 6.

Dimension 6		
Collocate	**Part of Speech**	**Loading**
large	Adjective	0.50
world	Noun	0.50
small	Adjective	0.47
society	Noun	0.45
city	Noun	0.43
u.s.	Noun	0.43
live	Verb	0.42
economy	Noun	0.41
many	Adjective	0.39
population	Noun	0.39
family	Noun	0.37
american	Adjective	0.36
community	Noun	0.35
other	Adjective	0.35

group	Noun	0.34
local	Adjective	0.34
country	Noun	0.33
state	Noun	0.32
most	Adjective	0.32
part	Noun	0.32
animal	Noun	0.30
organization	Noun	0.28
institution	Noun	0.27
body	Noun	0.27
grow	Verb	0.26
member	Noun	0.25
include	Verb	0.24
become	Verb	0.23
business	Noun	0.21

Table 16. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 6.

Dimension 6		
Node	**Part of Speech**	**Factor Scores**
large	Adjective	135.95
society	Noun	110.65
other	Adjective	107.25
many	Adjective	105.59
population	Noun	98.30
small	Adjective	87.49
state	Noun	85.43
group	Noun	84.45
country	Noun	79.48
people	Noun	78.40
live	Verb	72.80
family	Noun	69.58
community	Noun	64.95
such	Adjective	63.90
economy	Noun	63.74
part	Noun	63.43
local	Adjective	60.38
individual	Noun	59.83
world	Noun	59.66
social	Adjective	58.18
culture	Noun	57.81
include	Verb	57.67
different	Adjective	54.85
human	Adjective	54.33
political	Adjective	53.11

member	Noun	50.67
government	Noun	48.62
size	Noun	46.77
american	Adjective	44.84
organization	Noun	44.45
week	Noun	-4.44
decade	Noun	-4.44
end	Noun	-4.44
detail	Noun	-4.44
late	Adjective	-4.44
century	Noun	-4.44
beginning	Noun	-4.44
twentieth	Adjective	-4.44
let	Verb	-4.44
half	Noun	-4.44
night	Noun	-4.44
billion	Noun	-4.44
book	Noun	-4.44
nineteenth	Adjective	-4.44
run	Noun	-4.44
end	Verb	-4.44
distance	Noun	-4.44
interval	Noun	-4.44
middle	Noun	-4.44
career	Noun	-4.44
lag	Noun	-4.44
duration	Noun	-4.44
text	Noun	-4.44
minute	Noun	-4.44
turn	Verb	-4.44
exception	Noun	-4.44
name	Verb	-4.44
quarter	Noun	-4.44
recall	Noun	-4.44
mention	Verb	-4.44

The main topics enacted by the collocations that were formed by the collocates loading in Dimension 6 are related to society in broadest sense, both at a local (collocate examples: *city, community, local, state*) and a global level (collocate examples: *world, country, U.S.*), and members of society (collocate examples: *population, family, group, member, American*). Some of the collocational patterns that exemplify those topics are found in the following excerpts from the corpus.

*[27] Through deep involvement in an impoverished urban African **American community** called the Flats in a town outside Chicago, Stack uncovered residents' complex survival strategies based on extended kinship networks. (5_7_1.txt)*

*[28] Changes in employment probabilities for natives **living** in **cities** with **large** Mexican **populations** are much less related to **local** demand **conditions** than are changes in cities with few Mexicans. (5_6_127.txt)*

*[29] If these **foreign countries** are a small part of the **world economy**, then their fiscal change has a negligible impact on **other countries**. But if these **foreign countries** are a large part of the **world economy**, their increase in government purchases reduces world saving. (5_7_18.txt)*

*[30] The individual units of a foraging society, called bands, are usually made up of several related **nuclear families**, perhaps including some grandparents, siblings, or cousins. **Family relationships** are, thus, the basis for **social organization**. Food, for example, is distributed along lines of kinship. (5_7_3.txt)*

As with the cluster solution for Dimension 5, the cluster analysis for Dimension 6 revealed 3 clusters, from which only Clusters 2 and 3 contained nodes that formed collocations with the collocates from this dimension. In other words, although 1,188 nodes comprised Cluster 1, none of them are in the collocational networks identified in this dimension. All of the 386 nodes in Cluster 2 formed collocations with collocates present in this dimension. In addition, the 41 nodes found in Cluster 3, which coincided with the ones with the highest factor scores, also formed collocations.

The main categories identified for nodes in Cluster 3 reflect the same themes enacted by the collocates, namely: (1) local and global geography (e.g., *world, country, state, city, local, area, part*); (2) members of society (e.g., *population, group, people, family, community, individual, human, member, American, woman*); (3) aspects of society (e.g., *society, social, culture, economy, political, government, organization, market, life, business, live, become, include*); and (4) scale of intensity, amount or quality (e.g., *large, many, small, different, size, most, number, rate, major, other, such*). Most of the nodes in Cluster 2 belong to the same broad categories present in Cluster 3. Other categories that also emerged can be described as: (1) industry and economy (e.g., *company, work, industry, firm, corporation, corporate, production,*

produce, operate, growth, capital, risk, tax, spend, cost, income, money, stock, price); (2) research methods (e.g., *research, model, sample, variable, effect, control, scientific, evidence*); (3) anthropological and agricultural subjects (e.g., *specie, indigenous, ethnic, brain, hunt, primate, crop, anthropologist, wild*).

3.1.7. Collocational Dimension 7: Reporting Research

Dimension 7, Reporting Research, consists of 22 collocates, which are shown in Table 17. Overall, 658 unique collocations were formed across 322 unique nodes and the 22 collocates from this dimension. The 658 unique collocations represent the extended collocational network for the 22 collocates co-occurring across 322 nodes. Considering part of speech, 55% of the collocates are verbs, 32% are adjectives, and only 14% are nouns.

Table 18 shows the nodes with the highest positive and negative factor scores in this dimension. Of all the collocates in this dimension, 55% (12 out of 22) can be found in the top 100 nodes with the highest positive factor scores. More than half of them (7 of the 12; 58%) are among the 30 highest-ranking nodes. There were 1,293 nodes (80%) with negative factor scores, all with the same value of -2.95.

Table 17. Composition of Dimension 7.

Dimension 7		
Collocate	**Part of Speech**	**Loading**
suggest	Verb	0.66
empirical	Adjective	0.59
consistent	Adjective	0.53
support	Verb	0.52
recent	Adjective	0.48
indicate	Verb	0.47
show	Verb	0.45
previous	Adjective	0.42
examine	Verb	0.42
focus	Verb	0.38
base	Verb	0.36
research	Noun	0.34
present	Verb	0.34
finding	Noun	0.32
describe	Verb	0.30

early	Adjective	0.30
literature	Noun	0.26
report	Verb	0.25
genetic	Adjective	0.24
explain	Verb	0.22
come	Verb	0.20
future	Adjective	0.20

Table 18. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 7.

Dimension 7		
Node	**Part of Speech**	**Factor Scores**
study	Noun	139.18
research	Noun	130.30
finding	Noun	109.97
evidence	Noun	109.93
datum	Noun	95.99
result	Noun	83.86
theory	Noun	57.16
analysis	Noun	56.85
literature	Noun	56.52
work	Noun	49.26
suggest	Verb	49.00
model	Noun	48.73
section	Noun	48.03
chapter	Noun	46.72
recent	Adjective	46.52
previous	Adjective	46.25
estimate	Noun	43.70
empirical	Adjective	41.99
hypothesis	Noun	36.29
idea	Noun	35.06
article	Noun	34.63
difference	Noun	34.31
pattern	Noun	34.29
experience	Noun	34.10
participant	Noun	32.09
year	Noun	31.69
effect	Noun	31.13
demonstrate	Verb	29.62
social	Adjective	29.23
human	Adjective	28.15
week	Noun	-2.95
beginning	Noun	-2.95
half	Noun	-2.95

night	Noun	-2.95
billion	Noun	-2.95
book	Noun	-2.95
run	Noun	-2.95
end	Verb	-2.95
distance	Noun	-2.95
interval	Noun	-2.95
middle	Noun	-2.95
career	Noun	-2.95
lag	Noun	-2.95
duration	Noun	-2.95
text	Noun	-2.95
minute	Noun	-2.95
turn	Verb	-2.95
exception	Noun	-2.95
name	Verb	-2.95
quarter	Noun	-2.95
recall	Noun	-2.95
mention	Verb	-2.95
final	Adjective	-2.95
read	Verb	-2.95
recall	Verb	-2.95
phase	Noun	-2.95
million	Noun	-2.95
name	Noun	-2.95
seem	Verb	-2.95
date	Verb	-2.95

Similar to Dimension 5, the collocations in this dimension are part of the empirical research methods domain, but they are mostly used to report or describe research findings and literature review. As it can be seen from both tables, the collocates and nodes in this dimension are almost the reverse of what loaded into Dimension 5. That is, 58% of the verb collocates that loaded in Dimension 7 were high-scoring nodes in Dimension 5—*suggest, indicate, show, base, present, describe, report*—, and all of them can be used to report research findings. Moreover, more than half (52%) of the collocates from Dimension 5 appear in the top 50 nodes with the highest factor scores in Dimension 7. Therefore, it can be said that the extended collocational networks in both dimensions are shared across the same domains, with the emphasis in Dimension 7 being the act of reporting rather than conducting research and describing methods, as exemplified by the following excerpts.

*[31] Empirical **findings indicate** that an affective perceptual bias precedes ostensible changes in mood in response to treatment with serotonergic drug in major depressive disorder [74] (5_6_5.txt)*

*[32] Moreover, recent **literature suggests** that these accounts are not mutually exclusive (Danielmeier Ullsperger, 2011), i.e., the three **models** together could **explain** the behavioral slowing observed after an error. (5_6_90.txt)*

*[33] Thus, fairly **consistent evidence** from many studies indicates partner support in various forms is beneficial. (5_6_21.txt)*

*[34] **Results support** prior findings of a nonsignificant difference in effect size between the two treatments. They also **support findings** that treatment type accounts for only a very small proportion of outcome variance. However, the **results indicate** that some previously reported covariates of depression treatment outcome may be artifactual. (5_6_56.txt)*

The cluster analysis for Dimension 7 revealed a 3-cluster solution, with Cluster 1 comprised of 291 nodes, Cluster 2, 31 nodes, and Cluster 3, 1,293 nodes. As with other previous dimensions, one of the clusters—Cluster 3—contained none of the nodes that formed collocations in Dimension 7.

Therefore, the 322 nodes in Clusters 1 and 2 correspond to the unique nodes that formed collocations in this dimension. The nodes in Cluster 2 are the highest raking in terms of factor scores hence their categories are complimentary to the collocates that loaded in this dimension, that is: (1) literature (e.g., *literature, work, study, research, article*); (2) theory (e.g., *hypothesis, idea, theory, model*); (3) results and research methods (e.g., *finding, result, evidence, datum, analysis, effect, estimate, pattern, experience, suggest, demonstrate, empirical, consistent*); and (4) textual references (e.g., *section, chapter, recent, previous, year*). The nodes in Cluster 1 specify further the categories detailed for cluster 2, adding nodes related to study participants and instruments (e.g., *researcher, respondent, human, group, subject*) as well as the different types of research (e.g., *anthropological, ethnographic, clinical, organizational, environmental, psychological*) and knowledge areas in which research is being carried out (e.g., *economy, economics, childhood, history, family*).

3.1.8. Collocational Dimension 8: Individuals in Society

In total, 23 collocates loaded on Dimension 8, entitled Individuals in Society, comprising 849 unique collocations across 373 unique nodes. The 849 unique collocations represent the extended collocational network for the 23 collocates co-occurring across 373 nodes. As it can be seen on Table 19, there is a slight predominance of collocates as nouns (57%), with verbs and adjectives corresponding to 22% each.

The nodes with the highest positive factor scores are also from the same domains, as shown in the sample in Table 20. All 23 collocates (100%) can be found in the top 100 nodes with the highest positive factor scores, with almost two-thirds (65%) of them among the 30 highest-ranking nodes. Similar to other dimensions, all nodes with a negative factor score (1,242 nodes; 77%) had the same value (-3.44).

Table 19. Composition of Dimension 8.

Dimension 8		
Collocate	**Part of Speech**	**Loading**
woman	Noun	0.57
young	Adjective	0.52
old	Adjective	0.52
child	Noun	0.51
man	Noun	0.50
age	Noun	0.43
likely	Adjective	0.39
tend	Verb	0.37
poor	Adjective	0.37
education	Noun	0.33
work	Verb	0.32
home	Noun	0.32
more	Adjective	0.32
individual	Noun	0.30
gender	Noun	0.26
people	Noun	0.26
student	Noun	0.25
work	Noun	0.25
receive	Verb	0.24
school	Noun	0.24
find	Verb	0.23
life	Noun	0.22
compare	Verb	0.21

Table 20. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 8.

Dimension 8		
Node	**Part of Speech**	**Factor Scores**
woman	Noun	125.44
child	Noun	120.46
man	Noun	95.46
people	Noun	76.26
more	Adjective	71.78
age	Noun	63.43
old	Adjective	63.14
family	Noun	57.58
high	Adjective	52.44
live	Verb	51.12
individual	Noun	50.08
adult	Noun	47.26
young	Adjective	47.15
many	Adjective	47.10
other	Adjective	42.68
year	Noun	42.42
work	Verb	42.04
group	Noun	40.60
find	Verb	39.52
worker	Noun	39.12
work	Noun	38.86
health	Noun	37.50
school	Noun	36.58
home	Noun	36.42
student	Noun	35.63
society	Noun	33.84
study	Noun	33.34
percent	Noun	32.24
make	Verb	32.22
learn	Verb	32.10
beginning	Noun	-3.44
night	Noun	-3.44
book	Noun	-3.44
run	Noun	-3.44
end	Verb	-3.44
distance	Noun	-3.44
interval	Noun	-3.44
middle	Noun	-3.44
career	Noun	-3.44
lag	Noun	-3.44

duration	Noun	-3.44
text	Noun	-3.44
minute	Noun	-3.44
turn	Verb	-3.44
exception	Noun	-3.44
name	Verb	-3.44
quarter	Noun	-3.44
recall	Noun	-3.44
mention	Verb	-3.44
final	Adjective	-3.44
recall	Verb	-3.44
phase	Noun	-3.44
name	Noun	-3.44
date	Verb	-3.44
provider	Noun	-3.44
raw	Adjective	-3.44
aim	Verb	-3.44
internet	Noun	-3.44
connection	Noun	-3.44
computer	Noun	-3.44

The collocational patterns identified in this dimension are concerned with different members of society (e.g., *woman, child, man, individual, people, student*) and issues related to age, gender roles and place in society, education and work. The following excerpts show examples of how the collocations appear in the corpus.

*[35] Instead of looking for key biological factors to explain the social division of labor, a more useful strategy is to examine the **kinds** of **work** that **men** and **women** do in the context of specific societies to see how they relate to other cultural and historical factors. (5_7_7.txt)*

*[36] These different depictions of **men** and **women** reflect **gender roles**, societal beliefs about how **men** and **women** are expected to behave. (5_7_14.txt)*

*[37] **Single women** with <u>children</u> are probably more likely to live in poverty, less likely to have a **college education**, and more likely to **work** in service-sector **jobs**. (5_7_28.txt)*

*[38] As a result, the people in the twenty to sixty **age group** support a surprisingly large percentage of nonproductive **young** and **old people**. About 40 percent of the population in camps contributes little to the food supplies. This allocation of*

work to young and middle-aged adults allows for a relatively carefree childhood and adolescence and a relatively unstrenuous ***old age****. (5_7_2.txt)*

A 3-cluster solution was extracted for Dimension 8. In total, 1,242 nodes loaded in Cluster 1, 41 in Cluster 2, and 332 in Cluster 3. As with previous cluster analyses, none of the nodes in Cluster 1 formed collocations with the collocates in this dimension. Thus, the 373 nodes that formed collocations are found solely in Clusters 2 and 3.

An analysis of the possible node categories emerging from these clusters showed that the nodes in Cluster 2 appear to be specifications of the themes identified for the collocates and collocational patterns in this dimension, namely: (1) individuals in society (e.g., *adult, child, individual, man, people, student, woman, worker*; (2) the groups to which these individuals may belong (e.g., *family, community, group, class, job, school society, home, study, work, age, health, role*); (3) characteristics of individuals (e.g., *old, young, poor, different, likely*); (4) scales of quantity (e.g., *percent, many, more, high*); and (5) actions performed by individuals (e.g., *find, give, help, learn, live, make, receive, tend, work*). It can be said that the nodes in Cluster 3 further specify the categories described for Cluster 2, with the addition of a category that specifies time (e.g., *daily, early, everyday, prior, late, recent, century, day, hour, month, time*).

3.1.9. Collocational Dimension 9: Gaining Knowledge and Developing Concepts

Dimension 9, entitled Gaining Knowledge and Being Productive, is comprised of 18 collocates. Overall, there were 748 unique collocations across 385 unique nodes and the 18 collocates. The 748 unique collocations represent the extended collocational network for the 18 collocates co-occurring across 385 nodes. As it can be seen on Table 21, almost three-quarters (72%) of the collocates are verbs, with adjectives and nouns corresponding to 17% and 11%, respectively. Dimension 9 is the only one to have a collocate with a negative loading (*impact, -0.26*). Although present in Table 21, this collocate was not considered during the analysis.

Table 22 shows the nodes with the highest positive (mostly nouns) and negative factor scores. Eleven of the 18 collocates (61%) can be found in the top 100 nodes with the highest positive factor scores, and only 3 of the 18 collocates (17%) are in the top 30 highest-ranking nodes. Unlike other dimensions, the nodes with the highest negative factor scores (17 nodes) had differing values, albeit most of the remaining nodes had the same score of -2.72 (1,215 out of a total of 1,235 nodes with negative factor scores (76%)).

Table 21. Composition of Dimension 9.

Dimension 9		
Collocate	**Part of Speech**	**Loading**
learn	Verb	0.43
use	Verb	0.39
develop	Verb	0.38
new	Adjective	0.36
allow	Verb	0.35
give	Verb	0.33
require	Verb	0.31
need	Verb	0.31
available	Adjective	0.28
apply	Verb	0.28
get	Verb	0.26
share	Verb	0.25
information	Noun	0.25
make	Verb	0.23
use	Noun	0.22
produce	Verb	0.21
common	Adjective	0.21
take	Verb	0.20
(impact)	(Noun)	(-0.26)

Table 22. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 9.

Dimension 9		
Node	**Part of Speech**	**Factor Scores**
new	Adjective	74.06
more	Adjective	67.87
information	Noun	61.89
firm	Noun	51.85
approach	Noun	51.48
people	Noun	49.70
technology	Noun	46.43

method	Noun	40.19
resource	Noun	39.60
datum	Noun	39.45
idea	Noun	38.17
amount	Noun	37.46
technique	Noun	37.10
way	Noun	37.02
same	Adjective	36.58
make	Verb	34.61
language	Noun	32.59
time	Noun	32.45
money	Noun	32.41
system	Noun	32.15
many	Adjective	30.96
job	Noun	30.56
individual	Noun	30.49
food	Noun	29.50
knowledge	Noun	29.29
different	Adjective	29.28
measure	Noun	29.24
good	Adjective	28.65
tool	Noun	27.63
model	Noun	27.22
negative	Adjective	-12.60
assess	Verb	-12.52
examine	Verb	-11.83
environmental	Adjective	-11.52
potential	Adjective	-11.48
significant	Adjective	-11.37
direct	Adjective	-11.30
little	Adjective	-11.18
positive	Adjective	-11.08
economic	Adjective	-10.97
investigate	Verb	-10.85
major	Adjective	-10.72
health	Noun	-10.58
reduce	Verb	-10.51
strong	Adjective	-10.42
policy	Noun	-7.25
outcome	Noun	-5.07
beginning	Noun	-2.72
night	Noun	-2.72
book	Noun	-2.72
run	Noun	-2.72
end	Verb	-2.72

distance	Noun	-2.72
interval	Noun	-2.72
middle	Noun	-2.72
career	Noun	-2.72
lag	Noun	-2.72
duration	Noun	-2.72
text	Noun	-2.72
minute	Noun	-2.72

The main themes identified for the extended collocational networks that appear in this dimension seem to be related to gaining and developing knowledge, concepts and understanding, as the following excerpts exemplify.

[39] This type of 'reinforcement learning' algorithm [9] constitutes a powerful ***way*** *to* ***learn*** *about the environment and converges upon optimal behavioral policies (e.g., [61]). (5_6_5.txt)*

[40] The quadratic approximation ***method developed*** *here should be useful in many dynamic settings with asymmetric information (e.g., repeated partnerships and oligopoly games). (5_6_103.txt)*

[41] This macroeconomic ***model allows*** *us to study how the aggregate price level and the quantity of aggregate output are determined in the short run. (5_7_18.txt)*

[42] Apparently this differential learning history means that the words are likely to ***produce different*** *reports about observed events. (5_7_12.txt)*

[43] Conducting this type of ***analysis requires*** *identifiable changes in labor demand. (5_6_127.txt)*

The cluster analysis for Dimension 9 revealed a 3-cluster solution. All of the nodes from Clusters 2 and 3—totaling 327 and 56 nodes, respectively—formed collocational patterns with the collocates in this dimension. Out of the 1,232 nodes from Cluster 1, only two formed collocational patterns in this dimension.

The nodes with the highest factor scores in this dimension are distributed in both Clusters 2 and 3. The two nodes from Cluster 1—*policy* and *outcome*—have negative factor scores. The collocations formed by those two nodes that are present in the corpus are: *make * policy* and *produce * outcome*. The nodes in Cluster 2 can be divided into: (1) individuals (e.g., *people, researcher, individual, student*); (2)

information and resources (e.g., *datum, information, knowledge, source, resource, tool, material, thing*); (3) courses of action (e.g., *method, way, model, process, technique, approach, form, strategy, theory, system, learn, make, use*); (4) aspects of society (e.g., *government, law, culture, food, language, sense, technology, share*); (5) attribution of quality (e.g., *different, same, new, common, good, similar*); and (6) business (e.g., *money, company, firm, job, product, skill, opportunity*). In addition to further differentiating those categories, nodes in Cluster 3 also specify aspects of time (e.g., *year, day, period, month*) and statistics and research methods (e.g., *variable, treatment, test, statistic, statistical, regression, procedure, prediction, participant, hypothesis, inference, estimate, criterion, assumption*).

3.1.10. Collocational Dimension 10: Social Welfare

A total of 19 collocates loaded on Dimension 10, Social Welfare, comprising 638 unique collocations across 330 unique nodes. The 638 unique collocations represent the extended collocational network for the 19 collocates co-occurring across 330 nodes. As it can be seen on Table 23, there is a predominance of collocates as nouns (74%), followed by adjectives (21%) and verbs (4%).

As it can be seen in the sample in Table 24, the nodes with the highest positive factor scores are mostly nouns and adjectives. All 19 collocates (100%) can be found in the top 100 nodes with the highest positive factor scores, with 84% of them present in the top 30 highest-ranking nodes. A total of 1,285 nodes had the same negative factor score of -2.7 (80%).

Table 23. Composition of Dimension 10.

Dimension 10		
Collocate	**Part of Speech**	**Loading**
service	Noun	0.46
care	Noun	0.41
public	Adjective	0.40
provide	Verb	0.40
program	Noun	0.38
health	Noun	0.38
access	Noun	0.37
support	Noun	0.36

government	Noun	0.34
benefit	Noun	0.29
good	Adjective	0.28
financial	Adjective	0.27
resource	Noun	0.27
policy	Noun	0.26
opportunity	Noun	0.25
job	Noun	0.25
material	Noun	0.22
direct	Adjective	0.22
good	Noun	0.21

Table 24. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 10.

Dimension 10

Node	Part of Speech	Factor Scores
provide	Verb	127.07
health	Noun	70.69
service	Noun	69.30
public	Adjective	61.42
support	Noun	46.59
resource	Noun	46.41
access	Noun	46.08
education	Noun	45.52
government	Noun	45.46
cost	Noun	44.71
good	Adjective	44.60
receive	Verb	43.05
such	Adjective	42.86
care	Noun	39.14
policy	Noun	34.95
available	Adjective	34.07
program	Noun	34.03
offer	Verb	33.69
good	Noun	31.92
economic	Adjective	30.75
create	Verb	30.57
great	Adjective	29.85
financial	Adjective	29.44
job	Noun	29.23
social	Adjective	28.48
system	Noun	27.81
investment	Noun	27.59
sector	Noun	27.51

evidence	Noun	27.07
direct	Adjective	26.41
negative	Adjective	-2.70
assess	Verb	-2.70
examine	Verb	-2.70
environmental	Adjective	-2.70
significant	Adjective	-2.70
little	Adjective	-2.70
positive	Adjective	-2.70
investigate	Verb	-2.70
major	Adjective	-2.70
beginning	Noun	-2.70
night	Noun	-2.70
book	Noun	-2.70
run	Noun	-2.70
end	Verb	-2.70
distance	Noun	-2.70
interval	Noun	-2.70
middle	Noun	-2.70
career	Noun	-2.70
lag	Noun	-2.70
duration	Noun	-2.70
text	Noun	-2.70
minute	Noun	-2.70
turn	Verb	-2.70
exception	Noun	-2.70
name	Verb	-2.70
quarter	Noun	-2.70
recall	Noun	-2.70
mention	Verb	-2.70
recall	Verb	-2.70
phase	Noun	-2.70

As the name suggests, the main topic emerging from Dimension 10 is social welfare, in particular, access and provision of programs and benefits, as exemplified by the excerpts that follow.

*[44] The relationship between race and mental **health service** use is at least partly mediated by socioeconomic status, with some poor ethnic minority groups at greatest risk for low rates of treatment (Alegria et al., 2002). (5_6_64.txt)*

*[45] Indeed, in countries with generous welfare states, household **economic resources** and access to **medical care**, two of the mechanisms proposed to*

link incarceration and infant mortality (Wildeman, 2012b: 231–2), may be less affected by incarceration than in the United States, buffering the individual-level consequences of incarceration for infant mortality. (5_6_168.txt)

[46] They are poor, but unlike the nonworking poor, they are not eligible for many ***government supports*** *such as subsidized housing,* ***medical care*** *(Medicaid), and food stamps. […] Social Security also provides life* ***insurance benefits*** *to the survivors in cases of the death of a breadwinner and disability payments when a wage earner is unable to work. (5_7_16.txt)*

[47] Some, such as Greenpeace and Amnesty International, also wield influence in their ability to shape domestic and international politics by mobilizing ***public support*** *across the globe. (5_7_26.txt)*

The cluster analysis for Dimension 10 revealed a 3-cluster solution, with 1,285 nodes loaded into Cluster 1; 266 and 64 nodes into Clusters 2 and 3, respectively. The 330 unique nodes in this dimension are distributed in Clusters 2 and 3 only. Thus, none of the nodes from Cluster 1 appear in the collocational patterns identified for this dimension.

The nodes in Cluster 3 are the ones with the highest factor scores and seem to specify the themes identified in this dimension, namely as the following categories: (1) provision domains (e.g., *health, public, education, policy, economic, financial, social, insurance, medical*); (2) provision providers and recipients (e.g., *provider, institution, government, worker*); (3) provisions (e.g., *service, support, resource, access, care, sector, program, system, benefit, training, issue, goods*); (4) aspects that may govern provisions (e.g., *investment, evidence, cost, interest, control, impact, tax, decision, market, food, job, labor, firm, company*); (5) actions taken toward provisions (e.g., *provide, receive, offer, gain, create, reduce, produce, limit, improve, require, take, increase, pay, include, find help*); and (6) scale of quantity and quality (e.g., *good, great, new, direct, available, few, limited*). Nodes in Cluster 2 belong to the same broad categories identified for Cluster 3, specifying to greater detail each of them, particularly financial and economic aspects that are considered for provisions.

3.1.11. Collocational Dimension 11: Marking Time and Sectioning

Dimension 11, entitled Marking Time and Sectioning, consists of 15 collocates. Overall, there were 410 unique collocations across 225 unique nodes and the 15 collocates. The 410 unique collocations represent the extended collocational network for the 15 collocates co-occurring across 225 nodes. Table 25 presents the collocates that loaded in this dimension, with 40% being nouns, 33% adjectives, and 27% verbs.

Table 26 shows the nodes with the highest positive and negative factor scores. All 15 collocates (100%) can be found in the top 100 nodes with the highest positive factor scores, with 80% of them being in the top 30 highest-ranking nodes. Similar to the other dimensions the vast majority of nodes (86%) had negative factor scores with the same value. There were 1,390 nodes with the value of -1.95.

Table 25. Composition of Dimension 11.

Dimension 11		
Collocate	**Part of Speech**	**Loading**
next	Adjective	0.49
day	Noun	0.47
year	Noun	0.45
few	Adjective	0.42
period	Noun	0.41
chapter	Noun	0.37
first	Adjective	0.36
several	Adjective	0.35
century	Noun	0.32
spend	Verb	0.32
section	Noun	0.32
consider	Verb	0.24
begin	Verb	0.24
discuss	Verb	0.23
long	Adjective	0.21

Table 26. Sample of the 30 nodes with the highest positive and negative factor scores in Dimension 11.

Dimension 11		
Node	**Part of Speech**	**Factor Scores**
next	Adjective	78.63
year	Noun	63.73

last	Adjective	56.64
day	Noun	55.77
time	Noun	55.74
period	Noun	53.93
month	Noun	51.89
week	Noun	45.41
early	Adjective	45.19
few	Adjective	42.99
first	Adjective	41.51
decade	Noun	39.04
begin	Verb	38.30
previous	Adjective	36.99
more	Adjective	36.87
end	Noun	36.56
chapter	Noun	36.19
hour	Noun	35.36
consider	Verb	34.67
section	Noun	33.75
detail	Noun	32.24
past	Adjective	31.91
late	Adjective	30.00
century	Noun	27.48
study	Noun	27.35
examine	Verb	27.00
discuss	Verb	25.79
describe	Verb	22.25
see	Verb	22.19
beginning	Noun	21.08
negative	Adjective	-1.95
assess	Verb	-1.95
environmental	Adjective	-1.95
significant	Adjective	-1.95
positive	Adjective	-1.95
investigate	Verb	-1.95
academic	Adjective	-1.95
adult	Adjective	-1.95
advanced	Adjective	-1.95
ancient	Adjective	-1.95
civil	Adjective	-1.95
clean	Adjective	-1.95
cold	Adjective	-1.95
commercial	Adjective	-1.95
complete	Adjective	-1.95
deep	Adjective	-1.95
eastern	Adjective	-1.95

ethical	Adjective	-1.95
executive	Adjective	-1.95
exogenous	Adjective	-1.95
explicit	Adjective	-1.95
fast	Adjective	-1.95
former	Adjective	-1.95
fossil	Adjective	-1.95
fourth	Adjective	-1.95
german	Adjective	-1.95
identical	Adjective	-1.95
implicit	Adjective	-1.95
impossible	Adjective	-1.95
informal	Adjective	-1.95

Similar to Dimension 2, Dimension 11 seems to comprise collocations that, in addition to lexical structuring of discourse, also serve communicative functions related to the structure of the text. As the following excerpts show, there are two main domains that stand out: (1) collocational patterns that mark the passage of time; and (2) collocational patterns that mark the sectioning of text (for example, into chapters and sections).

*[48] Just as we did in our **previous chapter**, we can set up a basic spreadsheet to calculate the present values of the individual cash flows as follows. Notice that we have simply calculated the present values one at a time and added them up: What is the present value of 200 in one year, 400 the **next year**, 600 the **next year**, and 800 the **last year** if the discount rate is 12 percent? (5_7_17.txt)*

*[49] In this process, the introduction of the voluntary organization in a neighborhood initially has a negative effect on crime for a **period** of **time** (here it is from time points 0 to 5). (5_6_170.txt)*

*[50] This is not surprising based on the model we presented since, as we <u>discussed</u> in the **previous section**, not only are "Changers" large but also large firms tend to be "Changers." (5_6_144.txt)*

*[51] Recently things have changed. Over the <u>last</u> **several decades**, technological progress kept up its pace, but educational advancement slowed down. (5_7_18.txt)*

A 3-cluster solution was extracted for Dimension 11. As with most of the dimensions previously described, one of the clusters contained only nodes that did not form any collocational patterns with the collocates from this dimension. In this case, it was Cluster 3, with 1,390 nodes. Hence, Clusters 1 and 2 (27 and 198 nodes, respectively) contained the 225 nodes that are part of the collocational networks identified here.

The node categories identified in Cluster 1 can be described as: (1) markers of time (e.g., *year, day, time, period, month, week, decade, hour, century, early, late*); (2) markers of relative position in time or within the text (e.g., *next, previous, last, first, end, past*); (3) sections of text (e.g., *chapter, section*); (4) analyze and communicate (e.g., *consider, examine, discuss, detail, study, begin*); and (5) scale of quantity (e.g., *few, more*). In addition to further specifying the categories identified in Cluster 1, nodes in Cluster 2 can also be classified as: (1) courses of action and issues (e.g., *reason, factor, method, problem, issue, policy, topic, question, model, process, concept, research, way, characteristic, feature*); (2) descriptors of quality (e.g., *important, different, same, difference, new, critical*); (3) actions taken (e.g., *work, take, give, let, spend, occur, live, come, sell, emerge*); (4) intervals (e.g., *step, term, stage, distance, lot, run*); and (5) aspects and members of society (e.g., *person, generation, people, life, marriage, history, career, company, tradition, European*).

3.2. Cluster Analysis of the 11 Collocational Dimensions

As mentioned in the previous chapters, a hierarchical cluster analysis was carried out to identify categories according to which nodes can be grouped in terms of the collocational patterns identified by the dimensional collocational analysis. A 5-cluster solution was extracted. comprised of the following, which includes all the nodes:

1. Cluster 1: 241 nodes
2. Cluster 2: 76 nodes
3. Cluster 3: 1206 nodes
4. Cluster 4: 67 nodes
5. Cluster 5: 25 nodes

The description of each cluster is given in the subsections below.

3.2.1. Cluster 1

In Cluster 1, 125 of the 241 nodes (52%) have their highest factor scores in Dimensions 1, 3 or 4 (Finance and Economics, Social and Cultural Development, and Exerting Influence and Shaping Outcomes, respectively). The remaining nodes are spread out along the other dimensions, with the highest factor scores in Dimensions 2, 8, 10, and 11. The following are the main categories identified for the nodes in this cluster:

1. Individuals in society (*woman, child, man, worker, adult, student, participant, researcher, consumer*)
2. Development and learning (*understand, learning, information, shape, create, develop, affect, interest, find, approach, knowledge, growth, support, development, cognitive, skill, language*)
3. Political and socioeconomic aspects and systems (*health, service, system, public, government, technology, access, education, resource, financial, power, care, class, status, gender, biological*)
4. Business/Market (*product, production, productivity, money, profit, interest, share, sale, supply, demand, business, raise, stock, work, bank, produce, asset, company, sector*)
5. Time Period (*next, year, age, last, time, day, period, month, week, early, decade, hour, annual, century, late*)
6. Categorization (*particular, specific, various, certain, amount, number, size, several, relative, general, equal, degree, trait*)
7. Qualification (*important, same, more, less, few, old, young, new, great, good, basic, complex, major, key, common, simple, poor, easy, difficult, similar*)

Table 27 shows a sample of the 30 nodes that have the highest factor scores in this cluster, the part of speech they belong to, the dimensions in which they have positive factor scores, and the dimension in which they have the highest factor score. None of the top 100 nodes with high factor loadings comes from Dimension 7.

Table 27. Composition of Cluster 1.

Cluster 1			
Node	**Part of Speech**	**Dimensions**	**Highest Factor Scores**
woman	Noun	1, 3, 5-10	Dimension 8
child	Noun	1, 3, 5-10	Dimension 8
particular	Adjective	1-4, 6-8, 10	Dimension 2
man	Noun	3, 6, 8	Dimension 8
specific	Adjective	1-4, 7, 9	Dimension 2
understand	Verb	2-9, 11	Dimension 3
next	Adjective	4, 7, 11	Dimension 11
important	Adjective	2-7, 9-11	Dimension 2
worker	Noun	1, 6, 8-11	Dimension 1
new	Adjective	1-11	Dimension 9
various	Adjective	2, 6-7, 10-11	Dimension 2
certain	Adjective	2-4, 6-8	Dimension 2
learn	Verb	2-5, 8-11	Dimension 3
more	Adjective	1-11	Dimension 8
health	Noun	1-4, 6, 8, 10	Dimension 10
create	Verb	1-4, 6, 8-10	Dimension 3
service	Noun	2, 8-10	Dimension 10
production	Noun	1-3, 6, 10	Dimension 1
amount	Noun	1-2, 4-6, 8-11	Dimension 1
same	Adjective	1-9, 11	Dimension 2
raise	Verb	1-2, 8-9	Dimension 1
year	Noun	1-2, 4-9, 11	Dimension 11
age	Noun	1, 3-8, 11	Dimension 8
old	Adjective	1, 6, 8-9, 11	Dimension 8
receive	Verb	1, 5, 8-10	Dimension 1
system	Noun	1-4, 6-10	Dimension 3
information	Noun	1-10	Dimension 9
number	Noun	1-2, 4-9, 11	Dimension 1
public	Adjective	1-4, 6, 8, 10	Dimension 10
profit	Noun	1-2, 4, 9	Dimension 1

3.2.2. Cluster 2

The 76 nodes that comprised Cluster 2 are characterized by topics and themes identified in Dimensions 5 and 7, Empirical Research Methods and Reporting Research, respectively. The categories that emerged from the analysis are related to:

1. Statistics (*datum, effect, model, coefficient, regression, estimate, empirical, equation, correlation, parameter, error, statistical, experiment, statistic*)

2. Research (*study, research, theory, literature, work, idea, hypothesis, exist*)
3. Reporting (*show, indicate, present, report, suggest, describe, explain, demonstrate, illustrate, reveal, support, ask*)
4. Text (*section, chapter, recent, previous, first, table, column*)
5. Activity and mental verbs (*use, provide, include, base, analysis, estimate, obtain, apply, see, measure, study, test, conduct, examine, perform, observe, generate*)
6. Results (*finding, evidence, result, observation, conclusion*)
7. Effect (*negative, linear, significant, strong, consistent, positive, additional, multiple, main*)
8. Relations (*relationship, interaction, association*)

Table 28 shows a sample of the 30 nodes that have the highest factor scores in this cluster, the part of speech they belong to, the dimensions in which they have positive factor scores, and the dimension in which they have the highest factor score.

Table 28. Composition of Cluster 2.

Cluster 2

Node	Part of Speech	Dimensions	Highest Factor Scores
study	Noun	1, 3, 5-11	Dimension 7
use	Verb	1-10	Dimension 5
research	Noun	2-9, 11	Dimension 7
provide	Verb	2-5, 7-11	Dimension 10
finding	Noun	4-5, 7, 10-11	Dimension 7
evidence	Noun	2-3, 5-11	Dimension 7
show	Verb	1-8, 11	Dimension 5
datum	Noun	2-3, 5-11	Dimension 7
effect	Noun	1-11	Dimension 5
result	Noun	3-5, 7-11	Dimension 7
model	Noun	1-2, 4-11	Dimension 5
coefficient	Noun	5, 7	Dimension 5
include	Verb	1-11	Dimension 5
indicate	Verb	1, 5, 7	Dimension 5
base	Verb	1-3, 5-9	Dimension 5
present	Verb	1, 5, 7, 9-11	Dimension 5
theory	Noun	3-5, 7, 9	Dimension 7
analysis	Noun	2, 4-7, 9-11	Dimension 7
literature	Noun	6-7	Dimension 7
regression	Noun	5-7, 9	Dimension 5
report	Verb	1, 4-5, 7-8	Dimension 5

estimate	Verb	1, 5, 9	Dimension 5
work	Noun	1-3, 6-11	Dimension 7
suggest	Verb	3, 5, 7-8	Dimension 7
significant	Adjective	1, 3, 5-8	Dimension 5
section	Noun	4, 7, 10-11	Dimension 7
chapter	Noun	4, 7, 11	Dimension 7
recent	Adjective	3, 5-8, 11	Dimension 7
previous	Adjective	3, 5, 7, 11	Dimension 7
describe	Verb	2-3, 5, 7, 9, 11	Dimension 5

3.2.3. Cluster 3

Cluster 3 contained 1,206 nodes. Of those, 880 nodes had positive factor scores in one or more dimensions and 326 had negative factor scores in all dimensions. This cluster seems to comprise those nodes with the lowest positive factor scores across the dimensions. Some of the main categories identified from the nodes with highest factor scores are:

1. Evolution (*body, specie, human, animal, evolution, male, parent, evolutionary, history, future*)
2. Nation/Identity (*nation, European, minority, generation, Asian, foreign, African, global, national, native, resident*)
3. Qualification (*urban, modern, formal, hard, standard, distinctive*)
4. Literature (*article, review, paper, discussion, theoretical, word*)
5. Taking action (*spend, distinguish, keep, acquire, assess, go, leave, attend, buy, allow, let*)
6. Statistics and probability (*percentage, random, probability, distribution, ratio, yield, scale, data, predict*)

Table 29 shows a sample of the 30 nodes that have the highest factor scores in this cluster, the part of speech they belong to, the dimensions in which they have positive factor scores, and the dimension in which they have the highest factor score.

Table 29. Composition of Cluster 3.

Cluster 3			
Node	**Part of Speech**	**Dimensions**	**Highest Factor Score**
article	Noun	7	Dimension 7
body	Noun	1, 3, 5-7	Dimension 6
specie	Noun	6, 9	Dimension 6
nation	Noun	4, 6, 9	Dimension 6
history	Noun	3, 6-8, 11	Dimension 6
review	Noun	4, 7	Dimension 7
future	Adjective	1, 4-5, 7-8	Dimension 1
paper	Noun	4, 7	Dimension 7
college	Noun	8	Dimension 8
spend	Verb	1-2, 4, 6, 8, 11	Dimension 1
human	Noun	2-3, 6-8	Dimension 6
distinguish	Verb	2, 9	Dimension 2
line	Noun	2, 5, 7-9, 11	Dimension 7
discussion	Noun	4, 7, 11	Dimension 7
animal	Noun	6, 8	Dimension 6
beginning	Noun	11	Dimension 11
resident	Noun	6	Dimension 6
today	Noun	6	Dimension 6
theoretical	Adjective	3, 5, 7, 11	Dimension 7
keep	Verb	1-4, 9-10	Dimension 1
percentage	Noun	1, 4, 6	Dimension 1
random	Adjective	4-5	Dimension 5
acquire	Verb	3, 9-10	Dimension 3
strategy	Noun	3-4, 7, 9-10	Dimension 9
evolution	Noun	3-4	Dimension 3
national	Adjective	1, 6, 10	Dimension 6
assess	Verb	5	Dimension 5
relation	Noun	3, 8, 10	Dimension 3
allow	Verb	2-3, 5, 8-9	Dimension 8
fact	Noun	4, 7-8	Dimension 7

3.2.4. Cluster 4

Cluster 4 comprised 67 nodes, with all but one having the highest factor scores distributed among Dimensions 2, 3, 4 and 6 (Classification, Social and Cultural Development, Exerting Influence and Shaping Outcomes, and Local and Global Society, respectively). The exception was the verb *help*, which had the highest factor score in Dimension 7. The four nodes from Dimension 2 are the same ones with the

highest factor scores in that dimension (*other, different, such*, and *many*). They serve a more functional purpose rather than lexical in so much as they are relational classifiers functioning as attributive adjectives. The other node categories that emerged in this cluster reflect the topics and themes identified in those four dimensions, namely:

1. Features/attributes (*characteristic, aspect, condition, factor, feature, form, kind, pattern, thing, type, way, behavior*)
2. Aspects of life in society (*cultural, economic, political, social, culture, economy, environment, policy, performance, activity, live*)
3. Societal members/membership (*American, human, community, family, group, individual, member, organization, people, population, society, role*)
4. Courses of action and reasoning (*event, experience, issue, outcome, problem, question, situation, decision*)
5. Geographic area (*world, state, country, city, area, local*)
6. Qualification (*different, large, small*)
7. Change (*change, variable, become*)

Table 30 shows a sample of the 30 nodes that have the highest factor scores in this cluster, the part of speech they belong to, the dimensions in which they have positive factor scores, and the dimension in which they have the highest factor score.

Table 30. Composition of Cluster 4.

Cluster 4

Node	Part of Speech	Dimensions	Highest Factor Score
social	Adjective	1-10	Dimension 3
other	Adjective	1-11	Dimension 2
different	Adjective	1-9, 11	Dimension 2
large	Adjective	1-9	Dimension 6
behavior	Noun	1-9	Dimension 4
such	Adjective	1-10	Dimension 2
way	Noun	1-11	Dimension 4
society	Noun	2-4, 6, 8	Dimension 6
many	Adjective	1-6, 8-11	Dimension 2
cultural	Adjective	1-6, 8-9	Dimension 3
factor	Noun	1-7, 9, 11	Dimension 4
population	Noun	1, 3-6, 8, 10	Dimension 6
change	Noun	1-7, 9-10	Dimension 4
variable	Noun	1-2, 4-9	Dimension 4

small	Adjective	1-2, 4-6, 8	Dimension 6
state	Noun	1, 3-6, 10	Dimension 6
group	Noun	1, 3-9, 11	Dimension 6
culture	Noun	1-4, 6-10	Dimension 4
human	Adjective	1-8, 10	Dimension 2
country	Noun	1, 4, 6, 8, 9, 11	Dimension 6
people	Noun	1-11	Dimension 6
policy	Noun	1, 3-6, 9-11	Dimension 4
live	Verb	2-4, 6, 8, 11	Dimension 6
economic	Adjective	1-6, 9-10	Dimension 2
process	Noun	1, 3-4, 6-7, 9, 11	Dimension 4
family	Noun	1-3, 5-8, 10	Dimension 6
political	Adjective	2-4, 6, 10	Dimension 2
condition	Noun	1-6, 8-11	Dimension 4
community	Noun	4, 6, 8, 10	Dimension 6
economy	Noun	1, 4-10	Dimension 6

3.2.5. Cluster 5

Most of the nodes from Cluster 5 appear to belong to the same domain as the collocates in Dimension 1, Finance and Economics. They can be grouped into the following main categories:

1. Expression of quantity, magnitude or quality (*average, high, low, total, increase, level, percent, rate, value, reduce*)
2. Investments (*capital, cost, income, investment, return, risk, tax, value*)
3. Labor (*labor, wage, earn*)
4. Market (*firm, market, price, pay, tax*)

Table 31 shows the 25 nodes, the part of speech they belong to, the dimensions in which they have positive factor scores, and the dimension in which they have the highest factor score.

Table 31. Composition of Cluster 5.

Cluster 5

Node	Part of Speech	Dimensions	Highest Factor Score
average	Adjective	1, 5, 7-8, 11	Dimension 1
high	Adjective	1-4, 6-10	Dimension 1
low	Adjective	1, 3, 6, 8, 10-11	Dimension 1

total	Adjective	1-2, 4-6	Dimension 1
capital	Noun	1, 3, 6, 9-10	Dimension 1
cost	Noun	1-6, 8-10	Dimension 1
firm	Noun	1-6, 8-11	Dimension 1
income	Noun	1-2, 4-9, 11	Dimension 1
increase	Noun	1, 4-8	Dimension 1
investment	Noun	1, 6, 9-10	Dimension 1
labor	Noun	1-3, 8-10	Dimension 1
level	Noun	1-6, 8-9	Dimension 1
market	Noun	1-6, 9-10	Dimension 1
percent	Noun	1, 6-8, 11	Dimension 1
price	Noun	1-2, 4-7, 9-10	Dimension 1
rate	Noun	1-2, 4-11	Dimension 1
return	Noun	1, 7	Dimension 1
risk	Noun	1-2, 4-6, 9-10	Dimension 1
tax	Noun	1, 3, 5-6, 10	Dimension 1
value	Noun	1-10	Dimension 1
wage	Noun	1, 4-5, 8, 10	Dimension 1
earn	Verb	1, 8, 11	Dimension 1
increase	Verb	1-10	Dimension 1
pay	Verb	1-2, 8-11	Dimension 1
reduce	Verb	1-3, 5, 9-10	Dimension 1

In sum, the methodological approach developed in this study allowed the identification of 11 dimensions of extended collocational patterns, comprising a total of 10,474 unique collocations, as well as 5 clusters of nodes, which were then categorized.

This finalizes the description of the findings of this study. In the next chapter, the results are discussed, including conclusions reached, limitations, and possible future directions.

Chapter 4. Discussion and Conclusion

The results described above have shown that extended collocational networks realized as collocational dimensions do in fact exist for written academic texts. In each of the dimensions, it has been possible to identify sets of collocates that are shared across groupings of nodes and contribute to shape the topics and ideas presented in the text, showing that collocates do inter-collocate, that is, the collocates of one node collocate with other nodes. This answers the first research question posed in this study, “When considered across node words in a corpus, is it possible to identify collocational dimensions, that is, can the collocational networks associated with each individual node word be generalized to identify underlying ‘dimensions’: sets of collocates that tend to co-occur with a set of node words?”. In total, 11 dimensions of collocational patterns were identified for texts in the SBE corpus, namely: (1) Finance and Economics; (2) Classification; (3) Social and Cultural Development; (4) Exerting Influence and Shaping Outcomes; (5) Empirical Research Methods; (6) Local and Global Society; (7) Reporting Research; (8) Individuals in Society; (9) Gaining Knowledge and Developing Concepts; (10) Social Welfare; and (11) Marking Time and Sectioning.

Moreover, one of the gaps this study aimed to fulfill was to propose a methodological approach that would allow the identification of extended collocational patterns without having to first choose a single node as a starting point, as in the aforementioned studies (Williams, 1998, Brezina et al., 2015, Mollet et al., 2011). The results above show that the methodological approach developed for this study is valid and may possibly be used to identify collocational networks without the need of first choosing a node as a starting point of the analysis. While the methodological approach carried out was based on Berber Sardinha (2017), the goals of this study are different from the goal of analyzing collocations with respect to text and register. Still, the methods applied make use of the same statistical procedures. The interpretation goes beyond it to include semantic node categorization in terms of the dimensions identified in an effort to verify whether those collocational dimensions reflect the structuring of the lexicon in a written academic domain.

As mentioned in Chapter 2, the study design for this study should allow for characterization of the lexicon of written academic texts in the area of Social, Behavioral and Economic Sciences, its description and organization, in an attempt to

answer research questions 2 and 3, “Do those collocational dimensions reflect the aboutness of the lexicon in a written academic domain? To what extent can the collocational patterns in each dimension be interpreted semantically?” and “Do those collocational dimensions also reflect underlying functions in written academic texts?”. Methodologically, both the variables and the observations are words (collocates and nodes, respectively), thus the research questions and theoretical goals were aimed at discovering collocational patterns, structure and meaning relations in the lexicon represented by the corpus, not larger patterns of discourse. However, as it can be seen in the naming and interpretation of the dimensions, the collocational patterns enacted by the collocational dimensions have gone beyond the lexicon to clearly reflect discourse domains and show patterns of discourse. The research design as is should not have allowed this to be done. Nonetheless, it is a fact that the dimensions are reflecting discourse domains. That is not to say that the lexicon cannot be characterized, as seen by the interpretations carried out for the nodes with high factor scores. But the major patterns identified are reflective of discourse, not of the lexicon. Thus, it was possible to answer both research questions in terms of discourse.

A possible explanation for this may be the methodological choices made for the extraction of nodes and collocates. The cut-offs for dispersion, frequency, and strength of association established for this study were based on what has been described in the literature. As described in Chapter 2 and summarized in Table 3, for the nodes, only a frequency cut-off was established: the 2,000 most frequent nodes in the corpus. A combination of frequency, dispersion, and strength of association cut-offs was established for the collocates/collocations only. From the findings, it became apparent that the decision to not add a dispersion measure for the nodes may have played a large role in the nature of the findings in this study. For example, it could be the case that the nodes included are restricted to a few particular texts and/or to a single register or subdiscipline and hence the reason why the dimensions are more discipline-specific. Consequently, given the observations, the results show co-occurring collocational patterns that are more related to specific texts/register and specific subdisciplines. As an example, this can be seen clearly in Dimension 1, Finance and Economics, which is polarized to patterns from textbooks in the subdiscipline of Social and Economic Sciences. Another finding that corroborates this is the fact that the number of nodes with positive factor scores ranged from 14% to 28% only. In terms of aboutness, it can be said that Dimensions 1, 3, 4, 6, 8, 9, and 10 reflect aboutness of

discourse. As described in the previous chapter, all of those dimensions have collocational patterns that represent topics and subject-matters in areas subareas of Social, Behavioral and Economic Sciences. On the other hand, Dimensions 2 and 11, Classification and Marking Time and Sectioning, respectively, show collocational patterns that reflect underlying functions in discourse. The collocations seem to enact functional structuring of discourse and text cohesiveness, thereby functioning more as structuring of language rather than reflecting topics or aboutness. The collocates that loaded into Dimension 2 are mostly species nouns, known to occur in higher frequencies in academic registers. The nodes in this dimension are mostly relational classifiers functioning as attributive adjectives to the noun collocates. The collocational patterns identified in Dimension 11 also appear to serve communicative functions in discourse. The patterns mostly mark either the passage of time or relative positions within the text, as in text headings and subheadings. Dimensions 5 and 7 appear to be complementary, both showing collocational patterns associated with academic discourse, namely research methods and the reporting of research, fulfilling both lexical and functional communicative purposes. Therefore, it is possible to state that, in addition to aboutness, the collocational dimensions also reflect underlying functions in the corpus texts.

Since the collocational dimensions were shown to reflect patterns of discourse rather than lexicon and were strongly associated with the subdisciplines, it was found there was a need to further interpret the nature of the dimensions and attempt to characterize them with respect to register and subdiscipline, the two variables taken into consideration in the corpus design. Thus, for each dimension, descriptive statistics and analysis of variance were carried out based on the frequencies in the corpus of the 30 nodes with the highest positive factor scores and the 30 nodes with the highest negative factor scores. This resulted in a single score per text—similarly in a way to a ‘dimension score’ in MD studies. Unsurprisingly, Dimension 1 showed significant differences between the two subdisciplines (means: SES – 1494.13; BCS – 606.27 / p=.010; F=6.663). Not every dimension showed significant differences for subdiscipline by itself, although register contributed significantly to every dimension. At first glance, it might appear that register is driving the dimensions, more than subdisciplines. However, this finding should be interpreted with caution since it was based on frequency of nodes and there are higher frequency values in textbooks than articles simply due to text size. The estimated marginal means plots showed

interaction between subdisciplines mostly in terms of research articles. Dimensions 1, 3, 6, 8, 9, and 10 showed significant differences for the interaction between subdiscipline and register (Dimension 1: p=.000; F=95.135; R^2=.715 / Dimension 3: p=.000; F=21.845; R^2=.856 / Dimension 6: p=.005; F=8.045; R^2=.797 / Dimension 8: p=.049; F=3.934; R^2=.773 / Dimension 9: p=.001; F=11.554; R^2=.862 / Dimension 10: p=.000; F=65.763; R^2=.819). As mentioned, significant differences in registers may have been driven by the larger frequency occurrences in textbooks than articles. Moreover, there was an equal number of texts per discipline in the corpus.

A hierarchical cluster analysis was carried out on the factor scores of nodes across the 11 dimensions in order to answer the last research question: "Can scores on the collocational dimensions be used to cluster node words into underlying categories of words that behave in similar ways? If so, how can those clusters of node words be interpreted?". The cluster analysis revealed that nodes can be grouped into five different clusters in terms of collocational dimensions. The results show that clusters are a reflection of nodes that have high factor scores across multiple dimensions. Nodes in Cluster 1 were found to have the highest factor scores in Dimensions 1, 3 and 4, reflecting topics in Social and Economic Sciences. Nodes with the highest factor scores in Dimensions 5 and 7 were grouped into Cluster 2. Those two dimensions were found to be complementary, so it is not surprising to find their node words clustered together. Cluster 3 had the largest number of nodes and it appeared to concentrate the nodes with the lowest positive and negative factor scores across all dimensions. Nodes that grouped into Cluster 4 had the highest factor scores in Dimensions 2, 3, 4 and 6, though there was a higher proportion of nodes from Dimensions 4 and 6. Lastly, Cluster 5 contained nodes with the highest factor scores in Dimension 1 only. This aligns with the results of the collocational dimension and the limitation of node inclusion discussed above that shows Dimension 1 seems to be the most driven by subdiscipline. The second part of the research question asks how the interpretation of the clusters could be done. First, the interpretation took into account the descriptions done for the collocational dimensions in which the nodes occur with a higher score. Second, WordNet synsets and their descriptions were also used to arrive at node categories for each cluster.

Besides the lack of a dispersion measure for nodes, two other implications of the methods employed should be addressed. As mentioned previously, in this study the variables and the observations loaded into the factor analysis are linguistically the

same: nouns, verbs, and adjectives. This is highly unusual since the variable and unit of observation in a factor analysis are never the same. In a way, the word can't collocate with itself, but it can collocate with another word, which could be either a variable or an observation. This may explain the reason why, in almost all dimensions, most of the collocates with high loadings are the exact same words as the nodes with high factor scores. Indeed, in Dimensions 1, 9, 10 and 11, all the words are shared, that is, all the words entered into the factor analysis as collocates (variables) are the same as the nodes (observations) with the highest factor scores. It may also justify the fact that in some cases, a small set of nodes are driving the dimension. The other nodes also characterize the dimension, but the few nodes with the highest factor scores are really the ones responsible. This phenomenon may be compounded by the same word also being a collocate with high a loading. The dimensions with the lowest percentage of shared words are Dimensions 2 and 4, with 24% and 32% of collocates, respectively. Dimension 2 shows collocational patterns that appear to serve more lexicogrammatical functions rather than lexical structuring of discourse. In Dimension 4, the collocates seem to function as determinants of domains of influence and/or processes/contexts represented by the nodes.

Furthermore, as it can be seen from the results, with the exception of one collocate in Dimension 9, there were no collocates with negative loadings in any of the dimensions. This is something that has been previously reported in lexical multidimensional analysis studies, but without an explanation for why it may be happening. In this study, it might be due to the fact that nodes and collocates are the same word or it could be due to the lack of node dispersion. In other words, nodes may be from the same subdiscipline hence there is no positive or negative correlation, but rather groups of positive correlations that occur within the same subdiscipline because of node selection. That is, there is no negative counterpart. The investigation into the nature of the dimensions did show that both subdiscipline and register may be accountable for these results.

The lack of node dispersion may have been a major limitation since the nodes appear to be discipline-specific to a large extent. This could be reversed in future studies by having a wide dispersion of nodes across numbers of text, register and disciplines and/or subdisciplines so as to arrive at collocational patterns that are general and not discipline-specific. On the other hand, this may lead to the discovery

that a word might be well dispersed and still not have widely-dispersed collocational patterns.

In addition, the corpus collected for this study contains five different disciplines/research areas, with a total of 18 subdisciplines. At first, it was intended for the methodology to be carried out in all five NSF areas, however, time limitations and the desire for a more controlled proof-of-concept led to the use of texts from a single discipline area. Despite adding node dispersion, future directions may also be to conduct the analysis for the other four research areas, both per discipline and/or subdiscipline and an overall analysis encompassing the full corpus.

As mentioned in Chapter 2, there is a vast amount of NLP applications that make use of lexicon and semantic analysis to make decisions on how similar words are to each other and their distribution in multi-vector space. Those applications are widely used nowadays. Word2vec, for example, is considered the best tool available to produce word embeddings, but there is still no explanation as to why it is so successful. Most NLP applications have similar goals, but all of them abstract from individual words, making it even harder to find information on how and why they work so well. And it seems there is a desire to learn the reasons they work so well. Another NLP tool, RegisterExplorer[9] (Kutuzov et al., 2016) is an application that uses word embeddings trained on the British National Corpus (BNC) to identify semantic differences among words depending on register. An R-package, CorporaCoCo[10], developed by researchers from the Centre for Corpus Research, at the University of Birmingham, uses key collocates and key collocations to compare co-occurrence between two corpora, Although, from a completely different angle, this study may possibly start to shed light on the relationships among words and how they weave together to lexically structure discourse based on the collocational dimensions and the node words clustering together, especially with respect to discipline and register.

Another future direction might be to explore the relationship between discipline and phraseology. The collocational dimensions in this study have shown that collocational patterns tend to be discipline-specific and they can be interpreted both functionally and semantically. Thus, this methodology could be applied to larger spans of co-occurring words to investigate whether disciplines are equally phraseological and

[9] RegisterExplorer: English language registers explorer. Available at: http://vectors.nlpl.eu/explore/registers/
[10] CorporaCoCo: Corpora Co-Occurrence Comparison. Available at https://cran.r-project.org/web/packages/CorporaCoCo/index.html

how different discourse domains may vary in terms of phraseology, thereby adding to the phraseological body of research.

Finally, it is hoped that the results of this study may be used to inform and further develop the teaching and learning of English. The original motivation for this study came from teaching experiences in English for specific purposes and the lack of information and resources on collocational patterns beyond the famous collocation lists and dictionaries. Further research as well as development of teaching applications based on this and new studies are possibilities for continuing in this direction.

References

Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235–247.

Bartsch, S. (2004). *Structural and Functional Properties of Collocations in English*. Tübingen: Narr.

Berber Sardinha, T. (2004). *Linguística de Corpus*. Barueri, SP: Manole.

Berber Sardinha, T. (2017). Lexical priming and register variation. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical Priming: Applications and Advances*. John Benjamins: Amsterdam.

Berber Sardinha, T., Acunzo, C., & São Bento Ferreira, T. (2016). Dimensions of collocation in Brazilian Portuguese: Exploring the Brazilian Corpus on SketchEngine. In M. Diab & A. Villavicencio (Eds.), *Essays in Lexical Semantics in Honor of Adam Kilgarriff*. Springer: Berlin / New York.

Berber Sardinha, T., Ferreira, T., & Mayer, C. (in press). *A Dictionary of Portuguese Collocations*. London: Routledge.

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. (1993). Co-occurrence patterns among collocations: A tool for corpus-based lexical knowledge acquisition. *Computational Linguistics*, *19*, 549-556.

Biber, D. (1995). *Dimensions of Register Variation*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam/Philadelphia: John Benjamins.

Biber, D. (2009). Multi-dimensional approaches. In A. Lüedeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*. v.2. Berlin: Walter de Gruyter, pp. 822-855.

Biber, D. & Egbert, J. (2016). Register Variation on the Searchable Web: A Multi-Dimensional Analysis. *Journal of English Linguistics, 44*(2), 95-137. doi: 10.1177/0075424216628955.

Biber, D. & Jones, J. K. (2009). Quantitative Methods in Corpus Linguistics. In: A. Lüedeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*, v. 2. Berlin: Walter de Gruyter, pp. 1286-1304.

Biber, D. & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (2000). *Corpus Linguistics: Investigating Language Structure and Use*. (2nd ed.). Cambridge: Cambridge University Press.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and Writing in the University: A Multidimensional Comparison. *TESOL Quarterly*, *36*(1), 9-48. doi:10.2307/3588359.

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V. … (2004). *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. (ETS TOEFL Monograph Series, MS-25). Princeton, NJ: Educational Testing Service.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Brezina, V. (2018). *Statistics in Corpus Linguistics: a practical guide*. Cambridge: Cambridge University Press.

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics, 20*(2), 139-173.

Brinton, L., & Brinton, D. (2010). *The linguistic Structure of Modern English* (Revised ed.). Amsterdam: John Benjamins.

Cantos-Gomez, P. (2019). Multivariate Statistics Commonly Used in Multi-Dimensional Analysis. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multidimensional Analysis*. London: Bloomsbury Academic, pp.97-124.

Cantos-Gomez, P. (2013). *Statistical Methods in Language and Linguistic Research*. Sheffield: Equinox.

Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching*. Philadelphia: John Benjamins, pp. 67-88.

Conrad, S. (2011). Variation in corpora and its pedagogical implications. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on Corpus Linguistics*. Philadelphia: John Benjamins, pp.47-62.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2): 213-38.

Coxhead, A. (2013). Vocabulary and ESP. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes*. Boston: John Wiley & Sons, Inc, pp.115-135.

Crossley, S., Salsbury, T. & McNamara, D. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics, 36*(5), 570-590.

Desagulier. G. (2019). Can word vectors help corpus linguists?. *Studia Neophilologica*, Taylor Francis (Routledge): SSH Titles. https://doi.org/10.1080/00393274.2019.1616220. halshs-01657591v2

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, *28*(3), 157–169. https://doi.org/10.1016/j.esp.2009.02.002

Egbert, J. (2015). Sub-register and discipline variation in published academic writing: Investigating statistical interaction in corpus data. *International Journal of Corpus Linguistics*, *20*(1): 1-29.

Egbert, J. & Staples, S. (2019). Doing Multi-Dimensional Analysis in SPSS, SAS, and R. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multidimensional Analysis*. London: Bloomsbury Academic, pp.125-144.

Evert, S. (2005). *The Statistis of Word Coocurrence*: *Word Pairs and Collocations*. (Unpublished doctoral dissertation). Institut für maschinelle Sprachverarbeitung, Universität Stuttgart, Germany. Retrieved from: https://d-nb.info/976240033/34

Evert, S. (2008). Corpora and collocations. In: A. Lüedeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*, v. 2. Berlin: Walter de Gruyter, pp. 1212-1248.

Evert, S. (2017). *Making sense of multivariate analyses of linguistic variation*. Poster at the Corpus Linguistics 2017 Conference, Birmingham, UK. Retrieved from: http://www.stefan-evert.de/PUB/Evert2017CL/Evert2017CL_Poster.pdf

Evert, S. & Neumann, S. (2017). The impact of translation direction on characteristics of translated texts. A multivariate analysis for English and German. In G. De Sutter, M.-A. Lefer, and I. Delaere (eds.), *Empirical Translation Studies. New Theoretical and Methodological Traditions*. Mouton de Gruyter, Berlin.

Evert, S., Uhrig, P., Bartsch, S., & Proisl, T. (2017). E-VIEW-alation – a large-scale evaluation study of association measures for collocation identification. *Electronic lexicography in the 21st century. Proceedings of the eLex 2017 conference*. Leiden, The Netherlands, pp. 531–549.

Fellbaum, C. (ed.). (1998a). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fellbaum, C. (1998b). A Semantic Network of English: The Mother of All WordNets. *Computers and the Humanities, 32*, 209–220. https://doi.org/10.1023/A:1001181927857

Gablasova, D., Brezina, V. & McEnery, A.M. (2017). Collocations in corpus-based language learning research: identifying, comparing and interpreting the evidence. *Language Learning, 67*(S1): 155-79.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*, 305–327.

Goddard, C. & Schalley, A. (2010). Semantic Analysis. In N. Indurkhya, & F. J. Damerau (eds.), *Handbook of Natural Language Processing*, (2nd Ed.). Boca Raton: CRC Press, Taylor & Francis, 93-120.

Gorsuch, R. L. (2015), *Factor Analysis*, New York: Routledge.

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story* (Studies in Corpus Linguistics 71). Amsterdam and Philadelphia: John Benjamins.

Halliday, M. A. K. & Hasan, R. (1976). *Language, Context, and text: aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.

Hirst, G. (2009) Ontology and the Lexicon. In: Staab S., Studer R. (eds) *Handbook on Ontologies. International Handbooks on Information Systems.* Springer, Berlin, Heidelberg, p.269-292. https://doi.org/10.1007/978-3-540-24750-0_11

Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hoey, M. (2009). Corpus linguistics and word meaning. In: A. Lüedeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*, v. 2. Berlin: Walter de Gruyter, pp. 972-987.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hyland, K. (2015). Corpora and written academic English. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.

Hyland, K. & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly, 41*: 235-53.

Kutuzov, A., Kuzmenko, E., & Maraksova, A. (2016). Exploration of register-dependent lexical semantics using word embeddings. *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 26–34, Osaka, Japan, December 11-17.

Lei, L. & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, *(23)*2, Oct 2018, 216–243.

Levi, O., Goldberg, Y., & Dagan, I. (2015). Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225. https://doi.org/10.1162/tacl_a_00134

Loewen, S. & Gonulal, T. (2015). Exploratory Factor Analysis and Principal Components Analysis. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research*. New York: Routledge, pp.182-212.

McEnery, T. (2006). *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Abington, UK: Routledge.

McEnery, T. & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *CoRR*, abs/1301.3781. Retrieved from http://arxiv.org/abs/1301.3781.

Mollet, E., Wray, A., & Fitzpatrick, T. (2011). Accessing second-order collocation through lexical co-occurrence networks. In: T. Herbst, S. Faulhaber & P. Uhrig (Eds.), *Phraseological View of Language: A Tribute to John Sinclair*. Berlin: Mouton de Gruyter, pp. 87-122.

Pecina, P. (2010). Lexical association measures and collocation extraction. *Language Resources and Evaluation*, *44*(1-2), 137-158. https://doi.org/10.1007/s10579-009-9101-4.

Phillips, M. (1985). *Aspects of Text Structure: An Investigation of the Lexical Organisation of Text*. Amsterdam, Netherlands: North-Holland.

Phillips, M. (1989). *Lexical Structure of Text* [Discourse Analysis Monograph 12]. University of Birmingham, Birmingham, UK.

Poesio, M. (2000). Semantic analysis. In R. Dale, H. Moisl, and H. Somers (Eds.), *Handbook of Natural Language Processing,* pp. 93–122. New York, NY/Basel, Switzerland: Marcel Dekker.

Poudat, C. & Follette, P. (2012). Corpora and academic writing. In A. Boulton et al. (Eds.), *Corpus-informed Research and Learning in ESP: Issues and Applications*. Philadelphia: John Benjamins, pp. 167-191.

Prévot, L., Huang, C., Calzolari, N., Gangemi, A., Lenci, A., & Oltramari, A. (2010). *Ontology and the lexicon: A multidisciplinary perspective*. https://doi.org/10.1017/CBO9780511676536.002.

Rychly, P. (2008). A lexicographer-friendly association score. In P. Sojka & A. Horák (Eds.), *Proceedings of recent advances in Slavonic natural language processing, RASLAN 2008*. Brno: Masaryk University, pp. 6-9.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Simpson, R. C., Briggs, S. L., Ovens, J. & Swales, J. M. (2002). *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*, 487-512. https://doi.org/10.1093/applin/amp058

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. M. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive lexical semantics*, pp. 1-24. Amsterdam: John Benjamins.

Sinclair, J. M. (2004). *Trust the text: language, corpus and discourse*. London/New York/Canada: Routledge.

Sinclair, J., Jones, S., & Daley, R. (1970/2004). *English Collocation Studies: The OSTI Report*. London, UK: Continuum.

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, *33*(2), 149-183

Staples, S. & Biber, D. (2015). Cluster Analysis. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research*. New York: Routledge, pp.182-212.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of trouble with quantitative studies. *Functions of Language*, *2*(2), 23-56.

Stubbs, M. (1996). *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Stubbs, M. (2007). On texts, corpora and models of language. In Hoey, M., Mahlberg, M., Stubbs, M., & Teubert, W. (Eds.). *Text, Discourse and Corpora*. London: Continuum, pp. 127-161.

Szmrecsanyi, B., Biber, D., Egbert, J., & Franco, K. (2016). Towards more accountability: Modeling ternary genitive variation in Late Modern English. *Language Variation and Change, 28*(1), 1-29.

Williams, G. (1998). Collocational networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics*, *3*(1), 151–171. https://doi.org/10.1075/ijcl.3.1.07wil

Xiao, R. (2015). Collocation. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.

Xiao, R. & McEnery, T. (2006). Collocation, semantic prosody, and near synomyny: A corpus-linguistic perspective. *Applied Linguistics*, *27*(1), 103–129.

Zuppardi. M.C., & Berber Sardinha, T. A multi-dimensional view of collocations in academic writing. In Römer, Ute, Viviana Cortes & Eric Friginal (eds.). Forthcoming. *Advances in Corpus-based Research on Academic Writing. Effects of Discipline, Register, and Writer Expertise*. Amsterdam: John Benjamins.

Appendix 1 – Tree Tagger Tagset

ID	POS tag	Description	Example
1	CC	Coordinating conjunction	*and, but, or, &*
2	CD	Cardinal number	*1, four*
3	DT	Determiner	*the*
4	EX	Existential *there*	*there is*
5	FW	Foreign word	*et al.*
6	IN	Preposition or subordinating conjunction	*in, of, after*
7	JJ	Adjective	*nice*
8	JJR	Adjective, comparative	*nicer*
9	JJS	Adjective, superlative	*nicest*
10	LS	List item marker	*(1), A, ii*
11	MD	Modal	*will, can, could*
12	NN	Noun, singular or mass	*quality*
13	NNS	Noun, plural	*qualities*
14	NP	Proper noun, singular	*Jane*
15	NPS	Proper noun, plural	*Smiths*
16	PDT	Predeterminer	*such, all*
17	POS	Possessive ending	*'s*
18	PP	Personal pronoun	*I, he, she, it*
19	PP$	Possessive pronoun	*my, his, her, its*
20	RB	Adverb	*also, typically*
21	RBR	Adverb, comparative	*more, later*
22	RBS	Adverb, superlative	*most*
23	RP	Particle	*up, out*
24	SENT	End punctuation	*. ? !*
25	SYM	Symbol	*[ ] ( ) { }* + * =
26	TO	*to*	*to gain, to society*
27	UH	Interjection	*OK, oh, hum*
28	VB	Verb, base form	*learn, do*
29	VBD	Verb, past tense	*had, found*
30	VBG	Verb, gerund or present participle	*playing, weaving*
31	VBN	Verb, past participle	*begun, noted*
32	VBP	Verb, non-3rd person singular present	*are, provide*
33	VBZ	Verb, 3rd person singular present	*reveals, shows*
34	WDT	WH-determiner	*that, which*
35	WP	WH-pronoun	*who, what*
36	WP$	Possessive wh-pronoun	*whose*
37	WRB	WH-adverb	*where, how, when*
38	:	General joiner	*, ; / ' " - --*
39	$	Currency symbol	*$ € £ ¥*

Adapted from: Penn Treebank Tagset (http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/Penn-Treebank-Tagset.pdf)

Appendix 2 – Python Scripts

Script 01: Extract node and collocate window span (L4-R4) for target nodes (JNV), per corpus file. Filters: Only nouns, adjectives and verbs as nodes; sentence boundaries kept. Output: .csv with 4,058,524 rows.

```python
#----extract node-collocate window span for target nodes (JNV)----
import re, glob, os
import numpy as np
import sqlite3
import pandas as pd

#Define the window
window_n = 4

path = "corpus2\*.txt"

#Create a new directory if it doesn't exist
if not os.path.exists("01 output"):
    os.makedirs("01 output")

out_file = "01 output\\results.csv"
out_file_db = "01 output\\results.db"

record_complete_file = "01 output\\completed.txt"

#Target tags
pos_targets_ls = ["NN", "NNS", "JJ", "JJR", "JJS", "VB", "VBD", "VBG", "VBN", "VBP", "VBZ"]

#Setup headers
headers_s = """
var1
var2
var3
word_minus_4
word_minus_3
word_minus_2
word_minus_1
word
word_plus_1
word_plus_2
word_plus_3
word_plus_4
""".strip()
```

```python
# ----------------------------------
#Helpers

#This formats any line for proper csv layout, plus a line break
def formatCSVLine(ls_in):
    tmp_ls = [ ('"' + re.sub('"','_Q_',i) + '"' ) for i in ls_in]
    tmp_s = ",".join(tmp_ls) + "\n"
    return tmp_s

#This method gets the positions of the target tags in the list of tags in a file
def getPositions(array_a):
    tmp_a = np.array(array_a)
    positions_a = []
    for i in pos_targets_ls:
        positions_a += np.array( np.where(tmp_a == i) ).tolist()[0]
    positions_a.sort()
    return positions_a

#This gets added to the beginning and the end of each file, according to its length
dummy_ls = ["none"] * window_n

def writeToOutFile(filename_in, a_windows_combo_ls_in):

    #parse file name
    var1 = filename_in.split("_")[0]
    var2 = filename_in.split("_")[1]
    var3 = filename_in.split("_")[2]

    this_line_ls = [var1,var2,var3] + a_windows_combo_ls_in

    this_line_s = formatCSVLine(this_line_ls)
    open(out_file,"a").write(this_line_s)

def recordFile(file_in):
    # Assumes txt has been chopped off
    # Example: 4_5_67
    open(record_complete_file,"a").write( file_in + ".txt\n" )
```

```
78  def simplifyNJVTokenTags(array_a):
79      #This is an instance where a loop is needed because a decision has to be made at each
80      #element of the array, so we'll do the work here in a method
81      #The garbage collection will do its work so we can save memory
82      tmp_ls = []
83      for i in array_a:
84          if i[0] in ("N","J","V"):
85              tmp_ls.append(i[0])
86          else:
87              tmp_ls.append(i)
88      return tmp_ls
```

```
91  def filterForSENT(combo_a,window_n):
92      #This modifies windows like the following so that they get chopped
93      #off from SENT in either direction
94      #
95      #[',_,', 'J_many', 'SENT_.', 'N_neuron', 'V_respond', 'TO_to', 'N_cue', 'WDT_that', 'V_be']
96      #['none_none', 'none_none', 'none_none', 'N_neuron', 'V_respond', 'TO_to', 'N_cue', 'WDT_that', 'V_be']
97      #
98      #
99      #[',_,', 'SENT_.', 'N_da', 'N_neuron', 'V_respond', 'TO_to', 'N_cue', 'WDT_that', 'SENT_.']
100     #['none_none', 'none_none', 'N_da', 'N_neuron', 'V_respond', 'TO_to', 'N_cue', 'WDT_that', 'none_none']
101     #
102     #Then any analysis can ignore any variable none_none
103     #
104     #Do any of the combos begin with SENT?
105     tmp_ls = [i for i in range(0,len(combo_a)) if re.match("^SENT_.*$",combo_a[i])]
106     target_pos_ls = []
107     #this sets the positions for where the none combos should be placed
108     for i in tmp_ls:
109         if i < window_n:
110             target_pos_ls += list(range(0,i+1))
111         if i > window_n:
112             target_pos_ls += list(range(i,len(combo_a)))
113     #Modify the combo_a by chopping off at the begining and/or end from SENT on in
114     #either direction
115     for i in target_pos_ls:
116         combo_a[i] = "none_none"
117     return combo_a
```

```
119 def saveDfToSqlDb(df_in, path_out):
120
121     connection = sqlite3.connect(path_out)
122     cursor = connection.cursor()
123
124     cursor.execute("drop table if exists table1")
125     df_in.to_sql('table1', connection, if_exists='replace', index=False)
126
127     pd.read_sql("select * from table1", connection)
128
129     connection.commit()
130     connection.close()
131 # ------------------------------------
132
133 #Write first line to file
134 headers_ls = headers_s.splitlines()
135 open(out_file,"w").write( formatCSVLine( headers_ls ) )
```

```
137 for file in glob.glob(path):
138
139     #Get the name of the file to parse to numbers in the name
140     filename_s = os.path.basename(file)
141     filename_s = re.sub("\.txt$","",filename_s)
142     #Example output for filename_s:
143     #5_6_15.txt
144
145     #read in the text
146     text_s  = open(file,"r", errors="ignore").read()
147
148     #put each line in a ls
149     lines_ls = text_s.strip().splitlines()
150
151     #Get the tokens
152     tokens_ls = [i.split("\t")[2:4] for i in lines_ls if len(i.split("\t")) == 4]
153     pos_tokens_ls = [i[0] for i in tokens_ls]
154     lemma_tokens_ls = [i[1] for i in tokens_ls]
155     #Clean up the lemmas for any extra spaces and convert to lower case
156     lemma_tokens_ls = [i.lower().strip() for i in lemma_tokens_ls]
157
158     #Add the dummy list to the beginning and end of each
159     pos_tokens_ls = dummy_ls + pos_tokens_ls + dummy_ls
160     lemma_tokens_ls = dummy_ls + lemma_tokens_ls + dummy_ls
161
162     #Get the positions of each target token
163     target_pos_tokens_pn_ls = getPositions(pos_tokens_ls)
```

```
165     for i in target_pos_tokens_pn_ls:
166         start_n = i - window_n
167         end_n = (i + 1) + window_n
168         this_windows_tokens_ls = pos_tokens_ls[start_n:end_n]
169         this_windows_lemmas_ls = lemma_tokens_ls[start_n:end_n]
170
171         this_node_lemma = this_windows_lemmas_ls[window_n] #Array is zero based so no need to modify window_n value
172
173         #Let's not go any further if the node word isn't an English word
174         #Assumption: doesn't begin with an English letter
175         if not re.findall("^[^A-Za-z]", this_node_lemma):
176             #before combining the tokens and the lemmas, chop off any N, J, or V tag after the first letter
177             this_windows_tokens_ls = simplifyNJVTokenTags(this_windows_tokens_ls)
178
179
180             #Create the token_tag combo
181             this_windows_combo_ls = [t[0]+"_"+t[1] for t in list(zip(this_windows_tokens_ls,this_windows_lemmas_ls))]
182
183             #Chop off from SENT in either direction
184             this_windows_combo_ls = filterForSENT(this_windows_combo_ls,window_n)
185
186             # Example output for this_windows_combo_ls:
187             #[',_,', 'JJ_many', 'NP_da', 'NNS_neuron', 'V_respond', 'TO_to', 'NNS_cue', 'WDT_that', 'VBP_be']
188             writeToOutFile(filename_s, this_windows_combo_ls)
189
190             #print( filename_s, this_windows_combo_ls )
191
192     recordFile(filename_s)
193     print("Done: " + filename_s)
194
195 #Finally, let's save a copy to an sqlite db, just in case ---- gave an unexpected error
196 df = pd.read_csv(out_file, encoding = "ISO-8859-1")
197 saveDfToSqlDb( df, out_file_db)
198
199 print("Done")
```

Script 02: Create a wordlist per file and for the whole corpus. Outputs: .csv for lemmas per file (648,505 rows); .csv for lemmas aggregate (122,183 lemmas).

```
1   #----create wordlists per file and for the whole corpus----
2
3   import re, glob, os
4   import pandas as pd
5   from collections import Counter
6   from numba import jit
7   import sqlite3
8
9   path = "corpus2\*.txt"
10
11  #Create a new directory if it doesn't exist
12  if not os.path.exists("02 output"):
13      os.makedirs("02 output")
14
15  out_file_by_file = "02 output\\counts_tag_lemma_by_file.csv"
16  out_file_by_file_db = "02 output\\counts_tag_lemma_by_file.db"
17
18  out_file_aggregate = "02 output\\counts_tag_lemma_aggregate.csv"
19  out_file_aggregate_db = "02 output\\counts_tag_lemma_aggregate.db"
20
21  # ------------------------------------
22  #Helpers
23
24  #This formats any line for proper csv layout, plus a line break
25  def formatCSVLine(ls_in):
26      tmp_ls = [ ('"' + re.sub('"','_Q_',i) + '"' ) for i in ls_in]
27      tmp_s = ",".join(tmp_ls) + "\n"
28      return tmp_s
```

```
30  @jit(nopython=False) #This compiles methods with loops to c-like code, which speeds up processing
31  def writeToOutFile(filename_s,freqdist_ls):
32      for t in freqdist_ls:
33          ls = [filename_s, t[0], str(t[1])]
34          s = formatCSVLine(ls)
35          open(out_file_by_file,"a").write(s)
36
37  @jit(nopython=False)
38  def simplifyNJVTokenTags(array_a):
39      #This is an instance where a loop is needed because a decision has to be made at each
40      #element of the array, so we'll do the work here in a method
41      #The garbage collection will do its work so we can save memory
42      tmp_ls = []
43      for i in array_a:
44          if i[0] in ("N","J","V"):
45              tmp_ls.append(i[0])
46          else:
47              tmp_ls.append(i)
48      return tmp_ls
```

```
50  @jit(nopython=True)
51  def filterForSENT(combo_a,window_n):
52      #This modifies windows like the following so that they get chopped
53      #off from SENT in either direction
54      #
55      #[',_,', 'J_many', 'SENT_.', 'N_neuron', 'V_respond', 'TO_to', 'N_cue', 'WDT_that', 'V_be']
56      #['none_none', 'none_none', 'none_none', 'N_neuron', 'V_respond', 'TO_to', 'N_cue', 'WDT_that', 'V_be']
57      #
58      #
59      #[',_,', 'SENT_.', 'N_da', 'N_neuron', 'V_respond', 'TO_to', 'N_cue', 'WDT_that', 'SENT_.']
60      #['none_none', 'none_none', 'N_da', 'N_neuron', 'V_respond', 'TO_to', 'N_cue', 'WDT_that', 'none_none']
61      #
62      #Then any analysis can ignore any variable none_none
63      #
64      #Do any of the combos begin with SENT?
65      tmp_ls = [i for i in range(0,len(combo_a)) if re.match("^SENT_.*$",combo_a[i])]
66      target_pos_ls = []
67      #this sets the positions for where the none combos should be placed
68      for i in tmp_ls:
69          if i < window_n:
70              target_pos_ls += list(range(0,i+1))
71          if i > window_n:
72              target_pos_ls += list(range(i,len(combo_a)))
73      #Modify the combo_a by chopping off at the begining and/or end from SENT on in
74      #either direction
75      for i in target_pos_ls:
76          combo_a[i] = "none_none"
77      return combo_a
```

```
80  def saveDfToSqlDb(df_in, path_out):
81
82      connection = sqlite3.connect(path_out)
83      cursor = connection.cursor()
84
85      cursor.execute("drop table if exists table1")
86      df_in.to_sql('table1', connection, if_exists='replace', index=False)
87
88      pd.read_sql("select * from table1", connection)
89
90      connection.commit()
91      connection.close()
92
93  # ------------------------------------
94
95  headers_ls = ["file","tag_lemma","freq"]
96
97  open(out_file_by_file,"w").write(
98      formatCSVLine(headers_ls)
99  )
```

```
101 n=0
102 for file in glob.glob(path):
103
104     #Get the name of the file to parse to numbers in the name
105     filename_s = os.path.basename(file)
106     filename_s = re.sub("\.txt$","",filename_s)
107     #Example output for filename_s:
108     #5_6_15.txt
109
110     #read in the text
111     text_s  = open(file,"r", errors="ignore").read()
112
113     #put each line in a ls
114     lines_ls = text_s.strip().splitlines()
115
116     #Get the tokens
117     tokens_ls = [i.split("\t")[2:4] for i in lines_ls if len(i.split("\t")) == 4]
118
119     pos_tokens_ls = [i[0] for i in tokens_ls]
120     pos_tokens_ls = simplifyNJVTokenTags(pos_tokens_ls)
121
122     lemma_tokens_ls = [i[1] for i in tokens_ls]
123     #Clean up the lemmas for any extra spaces and convert to lower case
124     lemma_tokens_ls = [i.lower().strip() for i in lemma_tokens_ls]
125
126     combo_ls = [t[0]+"_"+t[1] for t in list(zip(pos_tokens_ls,lemma_tokens_ls))]
127
128     #Use Counter to get the Frequency distribution and all nodes
129     freqdist = Counter(combo_ls)
130     freqdist_ls = list( sorted( freqdist.items() ) )
131
132     writeToOutFile(filename_s,freqdist_ls)
133
134     n += 1
135     print(n, ": ", filename_s)
138 df = pd.read_csv(out_file_by_file, encoding = "ISO-8859-1")
139 saveDfToSqlDb(df, out_file_by_file_db)
140
141 #This provides the count for any item (node or collocate) in the entire corpus sample
142 pivot_df = df.groupby('tag_lemma')[['freq']].sum().reset_index()
143 pivot_df.to_csv(out_file_aggregate, index=False)
144 saveDfToSqlDb(pivot_df,out_file_aggregate_db)
145
146 print("Done")
```

Script 03: Get a wordlist of only nodes and their frequencies; get a wordlist of the 2000 most frequent nodes and their frequency counts; get a subset of the L4-R4 window span (from 01 results) for the top 2000 nodes. Filters: Only nouns, adjectives and verbs as nodes; 39 lemmas eliminated. Outputs: .csv with all node counts (68,575 nodes); node counts for 2,000 most frequent nodes; .csv subset (2,817,790 rows).

```python
#----get 2000 most frequent nodes, their L4-R4 window span, and their frequency counts----
# -*- coding: utf-8 -*-

import pandas as pd
from collections import Counter
import numpy as np
import sqlite3, os, re

node_sample_size_n = 2000

results_path = "01 output\\results.csv"

#Create a new directory if it doesn't exist
if not os.path.exists("03 output"):
    os.makedirs("03 output")

most_common_nodecounts_out_path = "03 output\\top_n_nodecounts.csv"
most_common_nodecounts_out_path_db = "03 output\\top_n_nodecounts.db"

results_subset_out_path = "03 output\\top_n_results_subset.csv"
results_subset_out_path_db = "03 output\\top_n_results_subset.db"

all_nodecounts_out_path = "03 output\\all_nodecounts.csv"
all_nodecounts_out_path_db = "03 output\\all_nodecounts.db"
```

```python
# --------- Helpers ---------------------

def saveDfToSqlDb(df_in, path_out):

    connection = sqlite3.connect(path_out)
    cursor = connection.cursor()

    cursor.execute("drop table if exists table1")
    df_in.to_sql('table1', connection, if_exists='replace', index=False)

    pd.read_sql("select * from table1", connection)

    connection.commit()
    connection.close()

# ----------------------------------------

#Put the column word (i.e., node) into a pandas dataframe
df = pd.read_csv(results_path, encoding = "ISO-8859-1")

#Now into a list
words_ls = list(df['word'])

#Use Counter to get the Frequency distribution of all nodes
freqdist = Counter(words_ls)

#Delete/filter nodes determined a priori
nodes_to_filter_ls = ['V_be', 'V_do', 'V_have', 'N_fig.', 'N_fl', 'N_fi', 'V_ai', 'V_not', 'J_not', 'N_not']

for i in freqdist:
    if ( re.match("^[NJV]_.$", i)) or (re.match("^[NJV]_n[^a-zA-Z]t$", i)):
        nodes_to_filter_ls.append(i)

for i in nodes_to_filter_ls:
    del freqdist[i]
```

```python
#-----Work on the most common ---------------

most_common_ls = freqdist.most_common( node_sample_size_n )
most_common_words_ls = [i[0] for i in most_common_ls]

#Store in a csv
most_common_nodecounts_df = pd.DataFrame(
            most_common_ls,
            columns=["node","freq"]
            )
most_common_nodecounts_df.to_csv(most_common_nodecounts_out_path,index=False)
saveDfToSqlDb(most_common_nodecounts_df, most_common_nodecounts_out_path_db)

#-----Create a subset of the results based only on the most common ---------------

top_words_indices_df = df[df['word'].isin(most_common_words_ls)]
top_words_indices_ls = top_words_indices_df.index.values.tolist()
results_subset_df = df[df.index.isin( top_words_indices_ls )]

results_subset_df.to_csv(results_subset_out_path,index=False)
saveDfToSqlDb(results_subset_df, results_subset_out_path_db)
```

```
85  #-----Record all node counts (just in case)-----------
86
87  all_nodecounts_ls = list( sorted( freqdist.items() ) )
88
89  #Store in a csv
90  all_nodecounts_df = pd.DataFrame(
91          all_nodecounts_ls,
92          columns=["node","freq"]
93          )
94
95  all_nodecounts_df.to_csv(all_nodecounts_out_path,index=False)
96  saveDfToSqlDb(all_nodecounts_df, all_nodecounts_out_path_db)
97
98  print("Done")
```

Script 04: Extract node and collocate frequencies per file. Outputs: 1 .csv per file.

```
1   #----node + collocate frequencies per file----
2   #----review 03 output to ensure all nodes + collocates are appropriate for further analysis----
3
4   # -*- coding: utf-8 -*-
5
6   import pandas as pd
7   from collections import Counter
8   import re, os
9
10  window_n = 4
11
12  #Create a new directory if it doesn't exist
13  if not os.path.exists("04 output"):
14      os.makedirs("04 output")
15
16  #Create a string to combine and split the node and the collocates
17  sep_s = "___TAB___"
18
19  #Build a list of the collocate columns
20  colocate_names_ls = ["word_minus_"+str(i+1) for i in range(0,window_n)]
21  colocate_names_ls += ["word_plus_"+str(i+1) for i in range(0,window_n)]
22
23  top_n_results_subset_path = "03 output\\top_n_results_subset.csv"
24  output_path = "04 output\\"
25
26  print("Setting up...")
```

```
30  #Read in the top N results
31  df = pd.read_csv(top_n_results_subset_path, encoding = "ISO-8859-1")
32
33  #Let's put in a df the unique var1, var2, var3 values
34  #We'll use this to separate the tabulations by file
35  vars_df = df[["var1","var2","var3"]]
36  vars_df = vars_df.drop_duplicates()
37  vars_df = vars_df.reset_index(drop=True)
```

```
40  for index, row in vars_df.iterrows():
41      #Rebuild the file name
42      var1, var2, var3 = list(row)
43      filename_s = "%s_%s_%s" % tuple(row)
44      print("Processing " + filename_s + "...")
45
46      #Let's focus on the results relevant to this file only
47      this_files_df = df[(df["var1"] == var1) & (df["var2"] == var2) & (df["var3"] == var3)]
48
49      #Build a list of  all the node+collocate combos
50      combos_ls = []
51      for index2, row2 in this_files_df.iterrows():
52          combos_ls += [ str(row2['word']) + sep_s + i for i in list( row2[colocate_names_ls] ) ]
53
54      #Build a FreQDist dictionary
55      freqdist_d = dict( Counter(combos_ls) )
```

```python
    #Format the freqdist_d to put into a df
    this_files_nodes_ls = [str(k) for k in freqdist_d.keys()]
    rows_ls = []
    for i in this_files_nodes_ls:
        try:
            node_collocate_ls = re.split(sep_s,i)
            d = {"file": filename_s,"node": node_collocate_ls[0], "collocate": node_collocate_ls[1], "freq": freqdist_d[i]}
            rows_ls.append(d)
        except:
            #Sometimes funky characters screw things up
            #So we'll just ignore them
            a=1

    #Create the df and save it
    out_df = pd.DataFrame(columns=["file","node","collocate","freq"])
    out_df = out_df.append(rows_ls,ignore_index=True)
    out_df.to_csv(output_path + filename_s + ".csv", index=False)

    print(filename_s, out_df.shape)

print("Done")
```


```python
    #Format the freqdist_d to put into a df
    this_files_nodes_ls = [str(k) for k in freqdist_d.keys()]
    rows_ls = []
    for i in this_files_nodes_ls:
        try:
            node_collocate_ls = re.split(sep_s,i)
            d = {"file": filename_s,"node": node_collocate_ls[0],
            "collocate": node_collocate_ls[1], "freq": freqdist_d[i]}
            rows_ls.append(d)
        except:
            #Sometimes funky characters screw things up
            #So we'll just ignore them
            a=1

    #Create the df and save it
    out_df = pd.DataFrame(columns=["file","node","collocate","freq"])
    out_df = out_df.append(rows_ls,ignore_index=True)
    out_df.to_csv(output_path + filename_s + ".csv", index=False)

    print(filename_s, out_df.shape)

print("Done")
```


Script 05: Merge script 04 files into a single .csv and create a node and collocate frequencies aggregate file (total frequencies). Outputs: .csv by file and aggregate.

```python
#----node + collocate frequencies per file in a single .csv
#and node + collocate frequencies aggregate file----
# -*- coding: utf-8 -*-

import pandas as pd
import re, os
import glob
import sqlite3

path_in = "04 output\*.csv"

#Create a new directory if it doesn't exist
if not os.path.exists("05 output"):
    os.makedirs("05 output")

path_out_byfile = "05 output\\node_collocates_by_file.csv"
path_out_byfile_db = "05 output\\node_collocates_by_file.db"

path_out_aggregate = "05 output\\node_collocates_aggregate.csv"
path_out_aggregate_db = "05 output\\node_collocates_aggregate.db"

headers_ls = ["file","node","collocate","freq"]
sep_s = "___TAB___"
```

```python
# --------- Helpers ----------------------

def saveDfToSqlDb(df_in, path_out):

    connection = sqlite3.connect(path_out)
    cursor = connection.cursor()

    cursor.execute("drop table if exists table1")
    df_in.to_sql('table1', connection, if_exists='replace', index=False)

    pd.read_sql("select * from table1", connection)

    connection.commit()
    connection.close()

# ----------------------------------------
# Work on the data to store the results by file

#Combine all the individual files into a single csv
n = 0
headers_s = ( ",".join(headers_ls) ) + "\n"
open(path_out_byfile,"w").write(headers_s)
for file in glob.glob(path_in):
    n += 1
    lines_ls = open(file,"r", errors="ignore").readlines()
    lines_ls = [i.strip() for i in lines_ls]
    lines_ls = lines_ls[1:]
    lines_s = "\n".join(lines_ls)
    open(path_out_byfile,"a").write(lines_s + "\n")
    print(n)
```

```python
#Place the combined data into a df
df = pd.read_csv(path_out_byfile, encoding = "ISO-8859-1", error_bad_lines=False)
saveDfToSqlDb(df, path_out_byfile_db)

# ----------------------------------------
# Work on the data to store the results as an aggregate

#Create a combined node + collocate column to group by
df['node_collocate'] = df['node'].map(str) + sep_s + df['collocate'].map(str)

#Calculate freq of the unique node + collocate values
aggregate_df = df.groupby('node_collocate')[['freq']].sum().reset_index()


#Rebuild the node and collocate columns
aggregate_df['node'] = aggregate_df.apply(lambda r: re.sub(sep_s+".+$","",r['node_collocate']),axis=1)
aggregate_df['collocate'] = aggregate_df.apply(lambda r: re.sub("^.+"+sep_s,"",r['node_collocate']),axis=1)
aggregate_df = aggregate_df.drop(columns = ['node_collocate'])
#Get rid of any collocates that ar dummy variables
aggregate_df = aggregate_df[aggregate_df['collocate'] != 'none_none']

#Save as csv
aggregate_df.to_csv(path_out_aggregate,index=False)
saveDfToSqlDb(aggregate_df, path_out_aggregate_db)

print("Done")
```

Script 06: Calculate text dispersion for target node + collocate pairs; eliminate collocates that are not JNV; get nodes + collocates that are JNV, their joint frequencies and file count. Filters: dispersion: 5% of the texts in the corpus (11.5 texts); only nouns, adjectives and verbs as collocates; 55 lemmas eliminated. Outputs: .csv with eliminated node + collocate pairs; .csv with filtered node + collocate pairs; .db with node + collocate pairs aggregate; .csv with node + collocate pairs frequency and file counts (50,575 node + collocate pairs after text dispersion cut-off).

```python
#----Calculate text file count for node+collocate pairs;
#Eliminate collocates that are not JNV and filter those that are JNV----
# -*- coding: utf-8 -*-

import pandas as pd
import sqlite3
import glob
import re, os
import numpy as np

print("Setting up...")
corpus_path = "corpus2\*.txt"
threshold_n = int(len(glob.glob(corpus_path)) * .05) #What's 5% of the corpus?

#Create a new directory if it doesn't exist
if not os.path.exists("06 output"):
    os.makedirs("06 output")

node_collocates_by_file_path_csv = "05 output\\node_collocates_by_file.csv"

#These are files for verifying the filtering process was accurate
node_collocate_filtered_path = "06 output\\node_collocate_filtered.csv"
node_collocate_eliminated_path = "06 output\\node_collocate_eliminated.csv"

node_collocates_aggregate_path_db = "06 output\\node_collocates_aggregate.db"
node_collocate_verbose_filecounts_path_csv = "06 output\\node_collocate_verbose_filecounts.csv"

sep_s = "__TAB__"

#These are the collocate tags to keep
filters_ls = ["N","J","V"]
```

```python
#These are the lemmas to eliminate form the collocate column (node column is already filtered)
collocate_lemmas_to_filter_ls = ['V_be', 'V_do', 'V_have', 'N_fig.', 'N_fl', 'N_fi', 'V_ai', 'V_not', 'J_not', 'N_not', 'N_s', 'N_i', 'J_s', 'V_s', 'N_b', 'N_n', 'N_n,t', 'N_x', 'N_f', 'N_t', 'N_p', 'N_e', 'N_r', 'N_u', 'N_k', 'N_y', 'N_l', 'N_h', 'N_c', 'N_j', 'N_m', 'N_w', 'N_o', 'N_z', 'N_v', 'N_g', 'N_d', 'N_a', 'N_q', 'N_@card@', 'N_et', 'N_de', 'N_ii', 'J_19th', 'J_20th', 'N_t.', 'N_1900s', 'N_1930s', 'N_1950s', 'N_1960s', 'N_1970s', 'N_1980s', 'N_1990s']


#Read in the data
df = pd.read_csv(node_collocates_by_file_path_csv, encoding = "ISO-8859-1")
```

```python
#-----Filter the df so that the collocates are only N, J, V------

#Make a copy of the df
copy_df = df.copy()

print("Deleting none_none...")
#Delete all none_none collocates, and reset the index b/c the filtering depends on the
#index position
copy_df = copy_df[np.logical_not(copy_df['collocate'] == "none_none")]
copy_df = copy_df.reset_index(drop=True)

print("Adding node_tag and collocate_tag columns...")
#Make two columns to hold just the node and collocate tags
#This process is sped up using numpy
func = np.vectorize( lambda i: re.sub("_.*","",i) )
copy_df['node_tag'] = func( np.array(copy_df['node']) )
copy_df['collocate_tag'] = func( np.array(copy_df['collocate']) )

print("Creating node_collocate column...")
#Make a column that combines the node and collocate values
copy_df['node_collocate'] = copy_df['node'] + sep_s + copy_df['collocate']

print("Creating node_collocate_tags column...")
#Make a column that combines the node and collocate values
copy_df['node_collocate_tags'] = copy_df['node_tag'] + "+" + copy_df['collocate_tag']

print("Creating list of indices to keep...")
#Put into a numpy array the indices of the filter items (i.e., to keep)
#and the indices of the collocate_lemmas_to_filter_ls
collocate_indices_a = np.array([])
```

```python
print(">> Selecting NJV...")
collocate_tags_a = np.array( copy_df['collocate_tag'] )
for i in filters_ls:
    collocate_indices_a = np.append(collocate_indices_a,np.where(collocate_tags_a == i)[0])

collocate_indices_lemmas_to_filter_a = np.array([])
collocates_a = np.array( copy_df['collocate'] )
for i in collocate_lemmas_to_filter_ls:
    collocate_indices_lemmas_to_filter_a = np.append(collocate_indices_lemmas_to_filter_a,np.where(collocates_a == i)[0])

print(">> Eliminating be, do, have...")
collocate_indices_a = np.setdiff1d( collocate_indices_a, collocate_indices_lemmas_to_filter_a)

print(">> Setting indices...")
#Now we have two sets of indices: ones to keep in the copy_df and ones to eliminate
collocate_indices_a = collocate_indices_a.astype(int)
collocate_indices_to_eliminate_a = np.setdiff1d( range(0, copy_df.shape[0]), collocate_indices_a)
collocate_indices_to_eliminate_a = collocate_indices_to_eliminate_a.astype(int)

#-----------------------------------------------------------------
#Create two dfs, one for the filtered copy_df version and one for the eliminated rows

print("Creating CSVs of rows to keep and rows to eliminate...")
node_collocate_filtered_df = copy_df.iloc[collocate_indices_a]
node_collocate_filtered_df = node_collocate_filtered_df.reset_index(drop=True)
node_collocate_filtered_df.to_csv(node_collocate_filtered_path,index=False)

node_collocate_eliminated_df = copy_df.iloc[collocate_indices_to_eliminate_a]
node_collocate_eliminated_df = node_collocate_eliminated_df.reset_index(drop=True)
node_collocate_eliminated_df.to_csv(node_collocate_eliminated_path,index=False)
```

```python
#Create the new database and write tha df to table1 pf the db

print("Creating database...")
conn = sqlite3.connect(node_collocates_aggregate_path_db)
node_collocate_filtered_df.to_sql("table1", conn, if_exists='append', index=False)
conn.commit()

#A cursor lets us modify the db (through the connection)
cur = conn.cursor()
#BREAK
#-----------------------------------------------------------------
#Create a view that shows us each node and collocate's file count and total frequency.

print("Creating node_collocate_verbose_filecounts view...")
sql = """
create view node_collocate_verbose_filecounts as
select node_collocate, node, collocate, node_tag, collocate_tag, node_collocate_tags,
sum(freq) as freq, count(file) as file_freq
from table1
group by node_collocate
having file_freq >= %s;
""".strip() % (threshold_n)
cur.execute(sql)
conn.commit()

print("Saving...")
#Let's export the node_collocate_verbose_filecounts view into a csv via pandas
node_collocate_verbose_filecounts_df = pd.read_sql("select * from node_collocate_verbose_filecounts;",conn)
node_collocate_verbose_filecounts_df.to_csv(
        node_collocate_verbose_filecounts_path_csv,
        index = False
        )
print("Done")
```

Script 07: Calculate logDice and filter node + collocate pairs according to logDice value. Filter: logDice > 7. Outputs: .csv for all results; .csv filtered for logDice cut-off; .csv with list of unique node + collocate pairs (17,893 unique collocations); .csvs with descriptive statistics for node and collocate tags.

```python
#----Calculate logdice for 2000 most frequent nodes and their collocates (only JNV)----
# -*- coding: utf-8 -*-
import pandas as pd
import math, os
import numpy as np

#Create a new directory if it doesn't exist
if not os.path.exists("07 output"):
    os.makedirs("07 output")
```

```
12 counts_tag_lemma_aggregate_df = pd.read_csv(
13         "02 output\\counts_tag_lemma_aggregate.csv", encoding = "ISO-8859-1"
14         )
15 node_collocate_verbose_filecounts_df = pd.read_csv(
16         "06 output\\node_collocate_verbose_filecounts.csv", encoding = "ISO-8859-1"
17         )
18  df = node_collocate_verbose_filecounts_df.copy()
19  df = df.reset_index(drop=True)
20
21  #-----Get rid of any rows that aren't cooperating
22  bad_rows_ls = [42256, 42286, 43959, 47657]
23 for i in bad_rows_ls:
24     df = df.drop(i, axis=0)
```

```
28 print("node_corpus_freq")
29 df['node_corpus_freq'] = df.apply(
30         lambda row: int(counts_tag_lemma_aggregate_df[counts_tag_lemma_aggregate_df['tag_lemma'] == row['node'] ]['freq']),
31         axis = 1
32         )
33
34 print("collocate_corpus_freq")
35 df['collocate_corpus_freq'] = df.apply(
36         lambda row: int(counts_tag_lemma_aggregate_df[counts_tag_lemma_aggregate_df['tag_lemma'] == row['collocate'] ]['freq']),
37         axis = 1
38         )
39
40 print("log_dice")
41 df['log_dice'] = df.apply(
42         lambda row: round( 14 + math.log2( ( (2 * row['freq'] ) / (row['node_corpus_freq'] + row['collocate_corpus_freq']) ) ), 4 ),
43         axis = 1
44         )
45
46 print("Writing")
```

```
50  df.to_csv("07 output\\all_results.csv", index=False)
51
52  df = df[df['log_dice'] >= 7.0]
53
54  df.to_csv("07 output\\all_results_logdice_filtered.csv", index=False)
55
56  #Finally, let's create some reports
57
58  #Just a dummy column for counting
59  df['N'] = 1
60  precision_n = 2
61
62  #How many individual node_collocates are there?
63 pd.pivot_table(
64     df,
65     index="node_collocate",
66     values="N",
67     aggfunc=len
68     ).to_csv("07 output\\report_uniq_node_collocates.csv")
```

```
70  #Stats
71 pd.pivot_table(
72         df,
73         index="node_tag",
74         values=["N","freq","log_dice"],
75         aggfunc={
76                 "N":len,
77                 "freq":np.sum,
78                 "log_dice":[np.mean,np.std,np.var]
79                 }
80         ).round(precision_n).to_csv("07 output\\report_node_tag.csv")
81
82 pd.pivot_table(
83         df,
84         index="collocate_tag",
85         values=["N","freq","log_dice"],
86         aggfunc={
87                 "N":len,
88                 "freq":np.sum,
89                 "log_dice":[np.mean,np.std,np.var]
90                 }
91         ).round(precision_n).to_csv("07 output\\report_collocate_tag.csv")
```

```
93  pd.pivot_table(
94          df,
95          index="node_collocate_tags",
96          values=["N","freq","log_dice"],
97          aggfunc={
98                  "N":len,
99                  "freq":np.sum,
100                 "log_dice":[np.mean,np.std,np.var]
101                 }
102         ).round(precision_n).to_csv("07 output\\report_node_collocate_tags.csv")
103
104 print("Done")
```

Script 08: Create matrix to be entered into multivariate statistical procedures. Output: .csv with 1,615 nodes as rows and 2,077 collocates as columns.

```
1   # -*- coding: utf-8 -*-
2   import pandas as pd
3   import numpy as np
4   import os
5
6   #Create a new directory if it doesn't exist
7   if not os.path.exists("08 output"):
8       os.makedirs("08 output")
9
10  all_results_logdice_filtered_df = pd.read_csv(
11          "07 output\\all_results_logdice_filtered.csv", encoding = "ISO-8859-1"
12          )
13  df = all_results_logdice_filtered_df.copy()
14  df = df[["node","collocate","log_dice"]]
15  df["log_dice"] = pd.to_numeric(df["log_dice"])
16
17  #Index in DataFrame.pivot_table() is what you want to convert to rows
18  df_count_pivot = df.pivot_table(
19          index=["node"],
20          columns = ["collocate"],
21          values = "log_dice",
22          aggfunc = len,
23          fill_value = 0,
24          margins=False
25          )
26  df_sum_pivot = df.pivot_table(
27          index="node",
28          columns = "collocate",
29          values = "log_dice",
30          aggfunc = np.sum,
31          fill_value = 0,
32          margins=False
33          )
34  #The rows are the nodes and the columns are the collocates
35  df_count_pivot.to_csv("08 output\\df_count_pivot.csv", index= True)
36  df_sum_pivot.to_csv("08 output\\df_sum_pivot.csv", index= True)
37  print("Done")
```

Script 09: Reduce the Script 08 matrix in order to meet requirements of PAF and/or PCA. Output: .csv with 1,615 nodes as rows and 323 collocates as columns.

```python
#create new 2 csv files, one for each node:collocate ratio (2:1 and 5:1)

import os, re
import pandas as pd

path_s = "08 output\\"

#Create a new directory if it doesn't exist
if not os.path.exists("09 output"):
    os.makedirs("09 output")

ratios_ls = ["2_1_ratio", "5_1_ratio"]
leftmost_columns_ls = ["node"]
collocates_21_ls = []
collocates_51_ls = []
data_ls = []
print("----Extracting lists of collocates----")
#extract lists of collocates from collocates_list.csv and
#append collocates from each ratio into a unique list
#make sure first row contains collocates for 2_1_ratio and
#second row for 5_1_ratio and that there are no extra commas in the file
file = "08 output\\collocate_lists.csv"
with open(file, encoding='utf-8-sig', errors='ignore') as file_in:
    #for each line, remove line endings, split at "," & append as list to a temporary list
    for line in file_in:
        data_line = line.rstrip().split(',')
        data_ls.append(data_line)

#for each list in "data_ls", remove the strings in the "ratios_ls" variable
for list in data_ls:
    #print(word)
    for w in ratios_ls:
        #print(w)
        if w in list:
            list.remove(w)
#save collocates from each nested list to a new variable according to ratio
collocates_21_ls = data_ls[0]
collocates_51_ls = data_ls[1]

#create column headings for each file according to ratio
ratio21_columns = leftmost_columns_ls + collocates_21_ls
ratio51_columns = leftmost_columns_ls + collocates_51_ls

#print("This is the 2_1_ratio collocates: " + str(collocates_21_ls))
#print("This is the 5_1_ratio collocates: " + str(collocates_51_ls))

print("----Reading Combo CSV file----")
files_ls = os.listdir(path_s)
#grab only combo_zeroall_df.csv and save it as a list in "files_ls"
files_ls = [i for i in files_ls if re.search("df_sum_pivot\.csv$",i) != None]
print(files_ls)

print("----Getting CSV to DF and appending to list----")
#Get the CSV file into a pandas data frame, save it as a df variable "pd_combo_df"
df_sum_pivot = pd.DataFrame()
for f in files_ls:
    df_sum_pivot = pd.read_csv(path_s + f)

print("----Creating 21_df and 51_df----")
#Create 2 pandas data frames with only columns in the "ratio21_columns" and "ratio51_columns" lists
ratio21_df = df_sum_pivot[ratio21_columns]
ratio51_df = df_sum_pivot[ratio51_columns]

print("----Printing final CSV output----")
#Output
ratio21_df.to_csv("09 output\\df_sum_pivot_ratio21.csv", index=False)
ratio51_df.to_csv("09 output\\df_sum_pivot_ratio51.csv", index=False)
print("Done")
```

Appendix 3 – List of the 1,615 nodes and 323 collocates entered into the multivariate exploratory factor analysis

Nodes (1,615)
J_able, J_absolute, J_academic, J_accurate, J_active, J_actual, J_adaptive, J_additional, J_adult, J_advanced, J_african, J_aggregate, J_aggressive, J_agricultural, J_alternative, J_american, J_ancient, J_animal, J_annual, J_anthropological, J_apparent, J_appropriate, J_asian, J_available, J_average, J_aware, J_bad, J_basic, J_behavioral, J_big, J_biological, J_black, J_broad, J_causal, J_central, J_certain, J_civil, J_classic, J_clean, J_clear, J_clinical, J_close, J_cognitive, J_cold, J_collective, J_commercial, J_common, J_comparative, J_competitive, J_complete, J_complex, J_considerable, J_consistent, J_constant, J_contemporary, J_continuous, J_corporate, J_correct, J_criminal, J_critical, J_cross-sectional, J_crucial, J_cultural, J_current, J_daily, J_deep, J_demographic, J_dependent, J_detailed, J_different, J_differential, J_difficult, J_direct, J_distinct, J_distinctive, J_diverse, J_domestic, J_dominant, J_dramatic, J_due, J_dynamic, J_early, J_eastern, J_easy, J_economic, J_educational, J_effective, J_efficient, J_emotional, J_empirical, J_english, J_enough, J_entire, J_environmental, J_equal, J_essential, J_estimated, J_ethical, J_ethnic, J_ethnographic, J_european, J_everyday, J_evolutionary, J_executive, J_existing, J_exogenous, J_expensive, J_experimental, J_explanatory, J_explicit, J_extensive, J_external, J_extreme, J_false, J_familiar, J_fast, J_federal, J_female, J_few, J_final, J_financial, J_first, J_following, J_foreign, J_formal, J_former, J_fossil, J_fourth, J_free, J_french, J_full, J_functional, J_fundamental, J_further, J_future, J_gay, J_general, J_genetic, J_geographic, J_german, J_global, J_good, J_great, J_hard, J_healthy, J_high, J_historical, J_huge, J_human, J_identical, J_illegal, J_immediate, J_implicit, J_important, J_impossible, J_independent, J_indian, J_indigenous, J_indirect, J_individual, J_industrial, J_informal, J_initial, J_institutional, J_interested, J_interesting, J_internal, J_international, J_japanese, J_key, J_large, J_last, J_late, J_least, J_legal, J_less, J_light, J_likely, J_limited, J_linear, J_little, J_local, J_long, J_long-term, J_low, J_main, J_major, J_male, J_many, J_marginal, J_married, J_mass, J_maximum, J_mean, J_medical, J_mental, J_middle, J_military, J_minimum, J_modern, J_monetary, J_moral, J_more, J_most, J_much, J_multiple, J_national, J_native, J_natural, J_necessary, J_negative, J_net, J_neural, J_new, J_next, J_nineteenth, J_normal, J_northern, J_nuclear, J_null, J_observed, J_obvious, J_old, J_ongoing, J_online, J_only, J_open, J_opposite, J_optimal, J_ordinary, J_organizational, J_original, J_other, J_outside, J_overall, J_own, J_particular, J_past, J_personal, J_physical, J_physiological, J_political, J_poor, J_popular, J_positive, J_possible, J_potential, J_powerful, J_practical, J_precise, J_present, J_previous, J_primary, J_prime, J_prior, J_private, J_productive, J_professional, J_prominent, J_psychological, J_public, J_racial, J_random, J_rapid, J_rational, J_raw, J_real, J_reasonable, J_recent, J_red, J_regional, J_regular, J_regulatory, J_related, J_relative, J_relevant, J_religious, J_reproductive, J_responsible, J_retail, J_rich, J_right, J_risky, J_robust, J_rural, J_same, J_sapiens, J_scarce, J_scientific, J_second, J_secondary, J_sensitive, J_sensory, J_separate, J_serial, J_serious, J_several, J_severe, J_sexual, J_sharp, J_short, J_short-term, J_significant, J_similar, J_simple, J_single, J_slow, J_small, J_social, J_southern, J_special, J_specific, J_stable, J_standard, J_statistical, J_steady, J_strategic, J_strong, J_structural, J_subjective, J_subsequent, J_substantial, J_successful, J_such, J_sufficient, J_superior, J_sure, J_surprising, J_systematic, J_technological, J_temporal, J_theoretical, J_third, J_top, J_total, J_traditional, J_tropical, J_true, J_twentieth, J_typical, J_unique, J_universal, J_unlikely, J_upper, J_urban, J_useful, J_valuable, J_various, J_vast, J_vertical, J_very, J_violent, J_visual, J_weak, J_wealthy, J_western, J_white, J_whole, J_wide, J_widespread, J_wild, J_willing, J_worth, J_wrong, J_young, N_ability, N_absence, N_abuse, N_access, N_account, N_achievement, N_act, N_action, N_activation, N_activity, N_actor, N_adaptation, N_adjustment, N_adult, N_advance, N_advantage, N_advertising, N_age, N_agency, N_agent, N_air, N_alcohol, N_alliance, N_allocation, N_alternative, N_amount, N_analysis, N_ancestor, N_animal, N_answer, N_anthropologist, N_anthropology, N_anxiety, N_anything, N_ape, N_application, N_approach, N_area, N_argument, N_arm, N_art, N_article, N_artifact, N_aspect, N_assessment, N_asset, N_association, N_assumption, N_attack, N_attempt, N_attention, N_attitude, N_attraction, N_attribute, N_author, N_authority, N_automobile, N_availability, N_average, N_awareness, N_baby, N_background, N_balance, N_band, N_bank, N_banking, N_barrier, N_base, N_basis, N_beginning, N_behavior, N_being, N_belief, N_benefit, N_bias, N_bill, N_billion, N_bird, N_birth, N_bit, N_black, N_blood,

N_board, N_body, N_bond, N_bone, N_book, N_border, N_bottom, N_boundary, N_boy, N_brain, N_branch, N_brand, N_brother, N_budget, N_building, N_burden, N_business, N_buyer, N_campaign, N_candidate, N_capacity, N_capital, N_car, N_carbon, N_card, N_care, N_career, N_case, N_cash, N_category, N_cause, N_cell, N_center, N_century, N_challenge, N_chance, N_change, N_chapter, N_characteristic, N_check, N_child, N_childhood, N_choice, N_chromosome, N_circumstance, N_citizen, N_city, N_civilization, N_claim, N_class, N_classification, N_climate, N_clothing, N_coefficient, N_cognition, N_colleague, N_collection, N_college, N_color, N_column, N_combination, N_communication, N_community, N_company, N_comparison, N_compensation, N_competition, N_complexity, N_component, N_computer, N_concentration, N_concept, N_concern, N_conclusion, N_condition, N_confidence, N_conflict, N_connection, N_consequence, N_consideration, N_constraint, N_construction, N_consumer, N_consumption, N_contact, N_context, N_contrast, N_contribution, N_control, N_corporation, N_correlation, N_cortex, N_cost, N_country, N_couple, N_course, N_court, N_credit, N_crime, N_crisis, N_criterion, N_crop, N_culture, N_curve, N_customer, N_cycle, N_damage, N_data, N_date, N_datum, N_daughter, N_day, N_deal, N_death, N_debate, N_debt, N_decade, N_decision, N_decline, N_decrease, N_deficit, N_definition, N_degree, N_delivery, N_demand, N_density, N_department, N_deposit, N_depression, N_descent, N_description, N_design, N_desire, N_detail, N_development, N_deviation, N_diet, N_difference, N_difficulty, N_dimension, N_direction, N_director, N_disadvantage, N_disaster, N_discipline, N_discount, N_discovery, N_discrimination, N_discussion, N_disease, N_disorder, N_disparity, N_display, N_distance, N_distinction, N_distribution, N_diversity, N_dividend, N_division, N_divorce, N_doctor, N_dollar, N_domain, N_dominance, N_driver, N_drug, N_duration, N_dynamics, N_earnings, N_earth, N_economics, N_economist, N_economy, N_edge, N_education, N_effect, N_efficiency, N_effort, N_elasticity, N_election, N_element, N_emission, N_emotion, N_emphasis, N_employee, N_employer, N_employment, N_end, N_energy, N_entry, N_environment, N_equality, N_equation, N_equilibrium, N_equity, N_era, N_error, N_estimate, N_estimation, N_ethnicity, N_evaluation, N_event, N_everyone, N_evidence, N_evolution, N_example, N_exception, N_exchange, N_executive, N_existence, N_expectation, N_expenditure, N_experience, N_experiment, N_explanation, N_export, N_exposure, N_expression, N_extent, N_eye, N_face, N_fact, N_factor, N_factory, N_failure, N_fall, N_family, N_farmer, N_father, N_favor, N_feature, N_fee, N_feedback, N_feeling, N_female, N_field, N_fieldwork, N_figure, N_finding, N_firm, N_fit, N_flow, N_focus, N_following, N_food, N_foot, N_force, N_forest, N_form, N_formation, N_fraction, N_framework, N_freedom, N_frequency, N_friend, N_fruit, N_fuel, N_function, N_fund, N_future, N_gain, N_game, N_gap, N_gender, N_gene, N_generation, N_girl, N_goal, N_gold, N_good, N_gorilla, N_government, N_grade, N_graph, N_ground, N_group, N_growth, N_habitat, N_half, N_hand, N_head, N_health, N_heart, N_height, N_help, N_heterogeneity, N_hierarchy, N_history, N_home, N_hospital, N_host, N_hour, N_house, N_household, N_housing, N_human, N_hundred, N_hunting, N_husband, N_hypothesis, N_ice, N_idea, N_identification, N_identity, N_illness, N_image, N_immigrant, N_impact, N_implication, N_import, N_importance, N_improvement, N_incarceration, N_incentive, N_income, N_increase, N_index, N_indicator, N_individual, N_industry, N_inequality, N_infant, N_inference, N_inflation, N_influence, N_information, N_innovation, N_input, N_insight, N_institution, N_instruction, N_instrument, N_insurance, N_intelligence, N_interaction, N_interest, N_internet, N_interpretation, N_interval, N_intervention, N_interview, N_introduction, N_investigation, N_investment, N_investor, N_island, N_isolation, N_issue, N_item, N_job, N_judge, N_judgment, N_justice, N_kind, N_knowledge, N_labor, N_laboratory, N_lack, N_lag, N_land, N_language, N_latter, N_law, N_leader, N_leadership, N_learning, N_length, N_letter, N_level, N_liability, N_life, N_light, N_likelihood, N_limit, N_limitation, N_line, N_link, N_list, N_literature, N_living, N_loan, N_lobe, N_location, N_logic, N_look, N_loss, N_lot, N_magnitude, N_majority, N_maker, N_male, N_man, N_management, N_manager, N_manipulation, N_manufacturing, N_margin, N_market, N_marriage, N_material, N_matter, N_mean, N_meaning, N_means, N_measure, N_measurement, N_meat, N_mechanism, N_medium, N_meeting, N_member, N_membership, N_memory, N_message, N_method, N_middle, N_migration, N_mile, N_million, N_mind, N_minority, N_minute, N_mobility, N_model, N_moderator, N_money, N_monkey, N_month, N_mood, N_mortality, N_mortgage, N_mother, N_motor, N_movement, N_movie, N_music, N_name, N_nation, N_nature, N_need, N_neighborhood, N_network, N_news, N_night, N_nonhuman, N_norm, N_note, N_nothing, N_notion, N_number, N_object, N_observation, N_offer, N_office, N_officer, N_official, N_offspring, N_oil, N_one, N_operation, N_opinion, N_opportunity, N_option, N_order, N_organism, N_organization, N_orientation, N_origin, N_other, N_outcome, N_output, N_owner,

N_page, N_pair, N_panel, N_paper, N_parameter, N_parent, N_part, N_participant, N_participation, N_partner, N_party, N_past, N_path, N_patient, N_pattern, N_pay, N_payment, N_people, N_percent, N_percentage, N_perception, N_performance, N_period, N_permission, N_person, N_personality, N_personnel, N_perspective, N_phase, N_phenomenon, N_phone, N_picture, N_piece, N_place, N_plan, N_plant, N_play, N_player, N_point, N_police, N_policy, N_politics, N_population, N_portfolio, N_portion, N_position, N_possibility, N_potential, N_poverty, N_power, N_practice, N_prediction, N_preference, N_premium, N_presence, N_present, N_president, N_pressure, N_price, N_primate, N_principle, N_prison, N_probability, N_problem, N_procedure, N_process, N_processing, N_producer, N_product, N_production, N_productivity, N_professional, N_professor, N_profit, N_program, N_progress, N_project, N_property, N_proportion, N_protection, N_provider, N_psychology, N_public, N_punishment, N_purchase, N_purpose, N_quality, N_quantity, N_quarter, N_question, N_race, N_radio, N_range, N_rate, N_rating, N_ratio, N_reaction, N_reader, N_reason, N_recall, N_recession, N_record, N_reduction, N_reference, N_reform, N_region, N_regression, N_regulation, N_relation, N_relationship, N_relative, N_religion, N_report, N_representation, N_requirement, N_research, N_researcher, N_resident, N_resource, N_respect, N_respondent, N_response, N_responsibility, N_rest, N_restriction, N_result, N_retrieval, N_return, N_revenue, N_review, N_reward, N_right, N_rise, N_risk, N_ritual, N_river, N_road, N_role, N_room, N_root, N_rule, N_run, N_safety, N_sale, N_sample, N_scale, N_scenario, N_scene, N_scholar, N_school, N_science, N_scientist, N_scope, N_score, N_search, N_season, N_section, N_sector, N_security, N_seed, N_segment, N_segregation, N_selection, N_seller, N_sense, N_sensitivity, N_sentence, N_sequence, N_series, N_service, N_session, N_set, N_setting, N_sex, N_shape, N_share, N_sheet, N_shift, N_shock, N_show, N_side, N_sign, N_signal, N_significance, N_similarity, N_site, N_situation, N_size, N_skill, N_skin, N_slope, N_society, N_solution, N_someone, N_something, N_son, N_sound, N_source, N_space, N_speaker, N_specie, N_species, N_specification, N_speech, N_spending, N_sport, N_spread, N_square, N_stage, N_standard, N_state, N_statement, N_statistic, N_status, N_step, N_stereotype, N_stimulus, N_stock, N_stone, N_store, N_story, N_strategy, N_street, N_strength, N_stress, N_structure, N_student, N_study, N_style, N_subject, N_subsidy, N_substitution, N_success, N_sum, N_summary, N_supply, N_support, N_surface, N_survey, N_survival, N_symbol, N_system, N_table, N_target, N_task, N_tax, N_teacher, N_team, N_technique, N_technology, N_television, N_temperature, N_tendency, N_term, N_test, N_testing, N_text, N_theme, N_theory, N_therapy, N_thing, N_thought, N_thousand, N_threat, N_tie, N_time, N_today, N_tool, N_top, N_topic, N_town, N_trade, N_trading, N_tradition, N_training, N_trait, N_transaction, N_transportation, N_treatment, N_tree, N_trend, N_trial, N_turn, N_tv, N_type, N_uncertainty, N_understanding, N_unemployment, N_union, N_unit, N_university, N_use, N_utility, N_value, N_variability, N_variable, N_variance, N_variation, N_variety, N_vehicle, N_version, N_victim, N_view, N_village, N_violence, N_voice, N_volatility, N_wage, N_war, N_water, N_way, N_wealth, N_week, N_weight, N_welfare, N_white, N_whole, N_wife, N_woman, N_word, N_work, N_worker, N_world, N_year, V_accept, V_accord, V_account, V_accumulate, V_achieve, V_acquire, V_act, V_adapt, V_add, V_address, V_adjust, V_adopt, V_affect, V_afford, V_agree, V_aim, V_allocate, V_allow, V_analyze, V_answer, V_appear, V_apply, V_argue, V_arise, V_arrive, V_ask, V_assess, V_assign, V_associate, V_assume, V_attempt, V_attend, V_attract, V_attribute, V_avoid, V_base, V_bear, V_become, V_begin, V_behave, V_believe, V_belong, V_bias, V_borrow, V_break, V_bring, V_build, V_buy, V_calculate, V_call, V_capture, V_care, V_carry, V_cause, V_challenge, V_change, V_characterize, V_charge, V_choose, V_classify, V_close, V_collect, V_combine, V_come, V_commit, V_communicate, V_compare, V_compete, V_complete, V_compute, V_concentrate, V_concern, V_conclude, V_conduct, V_confirm, V_connect, V_consider, V_consist, V_constitute, V_construct, V_consume, V_contain, V_continue, V_contribute, V_control, V_correlate, V_count, V_cover, V_create, V_cut, V_date, V_deal, V_decide, V_decline, V_decrease, V_define, V_demonstrate, V_denote, V_depend, V_depict, V_derive, V_describe, V_design, V_desire, V_detect, V_determine, V_develop, V_die, V_differ, V_direct, V_discuss, V_display, V_distinguish, V_distribute, V_divide, V_draw, V_drive, V_drop, V_earn, V_eat, V_eliminate, V_emerge, V_emphasize, V_employ, V_enable, V_encode, V_encounter, V_encourage, V_end, V_engage, V_enhance, V_ensure, V_enter, V_equal, V_establish, V_estimate, V_evaluate, V_evolve, V_examine, V_exceed, V_exclude, V_exhibit, V_exist, V_expand, V_expect, V_experience, V_explain, V_exploit, V_explore, V_expose, V_express, V_extend, V_extract, V_face, V_facilitate, V_fail, V_fall, V_favor, V_feed, V_feel, V_fight, V_fill, V_finance, V_find, V_fit, V_fix, V_focus, V_follow, V_forage, V_form, V_gain, V_gather, V_generalize, V_generate, V_get, V_give,

V_go, V_govern, V_grant, V_grow, V_guide, V_happen, V_hear, V_help, V_highlight, V_hire, V_hold, V_hunt, V_identify, V_ignore, V_illustrate, V_imagine, V_implement, V_imply, V_impose, V_improve, V_include, V_increase, V_indicate, V_influence, V_integrate, V_interact, V_interpret, V_introduce, V_invest, V_investigate, V_involve, V_issue, V_join, V_keep, V_kill, V_know, V_lack, V_lead, V_learn, V_leave, V_lend, V_let, V_lie, V_like, V_limit, V_link, V_list, V_live, V_locate, V_look, V_lose, V_lower, V_maintain, V_make, V_manage, V_marry, V_match, V_maximize, V_mean, V_measure, V_meet, V_mention, V_minimize, V_miss, V_model, V_moderate, V_monitor, V_motivate, V_move, V_name, V_need, V_note, V_observe, V_obtain, V_occupy, V_occur, V_offer, V_open, V_operate, V_organize, V_own, V_participate, V_pass, V_pay, V_perceive, V_perform, V_place, V_play, V_pose, V_possess, V_predict, V_prefer, V_prepare, V_present, V_press, V_prevent, V_process, V_produce, V_promote, V_propose, V_protect, V_prove, V_provide, V_publish, V_purchase, V_pursue, V_put, V_raise, V_range, V_reach, V_read, V_realize, V_recall, V_receive, V_recognize, V_reduce, V_refer, V_reflect, V_regard, V_reject, V_relate, V_rely, V_remain, V_remember, V_repeat, V_replace, V_report, V_represent, V_reproduce, V_require, V_respond, V_restrict, V_result, V_return, V_reveal, V_review, V_rise, V_rule, V_run, V_satisfy, V_save, V_say, V_see, V_seek, V_seem, V_select, V_sell, V_send, V_serve, V_set, V_shape, V_share, V_shift, V_show, V_solve, V_speak, V_specify, V_spend, V_stand, V_start, V_state, V_stay, V_store, V_study, V_suffer, V_suggest, V_summarize, V_supply, V_support, V_suppose, V_surround, V_survive, V_take, V_teach, V_tell, V_tend, V_test, V_think, V_trade, V_train, V_travel, V_treat, V_try, V_turn, V_underlie, V_understand, V_use, V_vary, V_wait, V_want, V_watch, V_wear, V_win, V_work, V_write, V_yield

Collocates (323)

J_social, J_different, J_other, J_new, N_people, J_such, N_rate, J_many, N_research, N_firm, J_high, N_effect, J_human, V_use, N_group, N_study, V_make, N_price, V_increase, N_change, J_more, N_model, N_market, N_value, V_provide, J_large, N_information, N_behavior, J_good, N_difference, N_way, N_level, V_include, N_time, N_datum, J_economic, N_variable, N_child, N_system, J_great, N_cost, N_increase, J_cultural, N_culture, N_policy, N_factor, N_evidence, N_population, N_number, V_reduce, J_same, J_important, V_give, N_life, N_resource, N_country, N_individual, N_income, N_woman, N_amount, N_society, V_understand, N_process, N_health, J_low, N_analysis, V_find, N_government, J_political, V_show, N_problem, N_pattern, V_base, N_year, N_size, N_percent, N_theory, J_small, N_result, N_relationship, N_work, N_world, V_produce, N_job, N_condition, V_pay, N_family, V_live, V_affect, V_take, N_role, N_question, N_finding, N_power, J_particular, V_involve, J_specific, V_require, N_food, N_money, N_worker, N_activity, N_labor, N_age, N_student, N_method, N_language, N_thing, N_term, V_receive, J_first, V_learn, N_outcome, N_risk, V_develop, V_become, V_lead, N_wage, N_issue, V_work, V_help, N_researcher, N_ability, N_chapter, N_sample, J_significant, N_state, N_estimate, N_form, N_response, N_business, V_create, N_production, N_product, N_environment, N_area, V_depend, N_tax, J_total, N_table, N_performance, N_interest, N_man, N_development, N_structure, J_own, J_similar, V_need, N_section, N_decision, J_average, N_knowledge, N_measure, J_early, N_participant, J_few, N_part, N_literature, N_experience, N_set, N_approach, V_change, N_school, N_investment, J_local, J_several, V_suggest, V_associate, J_most, J_public, N_growth, N_demand, N_period, J_recent, N_source, N_control, N_use, N_company, V_get, V_focus, J_negative, N_capital, N_service, V_report, V_consider, V_study, N_idea, V_measure, J_positive, N_task, N_word, N_economy, V_present, J_major, N_practice, V_see, N_member, N_opportunity, N_point, J_strong, N_community, V_examine, N_organization, V_determine, N_good, N_gender, J_key, N_skill, V_explain, N_type, V_describe, N_characteristic, V_follow, N_hypothesis, V_raise, N_experiment, J_financial, N_technology, N_rule, N_event, N_category, V_influence, N_benefit, V_vary, V_identify, N_attention, J_next, J_future, N_variation, N_support, N_day, N_law, N_home, N_action, V_expect, J_simple, V_grow, N_impact, N_place, J_various, N_reason, J_long, N_test, N_function, N_aspect, N_feature, N_program, J_complex, J_less, J_basic, V_indicate, V_come, J_natural, J_physical, J_cognitive, J_poor, N_example, J_certain, V_apply, V_reflect, V_know, V_compare, V_allow, N_bank, V_spend, N_profit, N_education, V_support, N_status, J_common, V_discuss, N_strategy, V_occur, V_share, J_likely, V_represent, V_begin, N_u.s., J_american, N_body, N_force, J_empirical, J_available, V_hold, V_improve, J_consistent, J_individual, V_maintain, V_tend, N_century, N_return, N_care, J_sexual, N_race, J_old, N_animal, N_equation, N_belief, J_previous, N_context, N_institution, J_second, N_line, N_kind, V_go, N_asset, J_young, N_regression, N_error,

N_brain, V_estimate, J_genetic, V_perform, N_access, N_human, J_direct, V_cause, N_city, N_share, J_relative, N_treatment, N_material, V_shape

Nodes and collocates are preceded by their part-of-speech. J=Adjective; N=Noun; V=Verb

Appendix 4 – Numeric expressions used to calculate factor scores

Factor 1

(ZN_wage + ZJ_low + ZN_price + ZN_rate + ZV_increase + ZJ_average + ZJ_high + ZN_increase + ZN_income + ZN_investment + ZN_capital + ZN_cost + ZN_level + ZN_return + ZN_percent + ZV_reduce + ZN_profit + ZN_share + ZV_expect + ZN_firm + ZJ_total + ZN_risk + ZN_tax + ZN_production + ZN_demand + ZN_growth + ZV_pay + ZN_market + ZN_interest + ZN_labor + ZV_associate + ZJ_less + ZN_amount + ZN_worker + ZN_value + ZJ_relative + ZV_raise + ZV_lead + ZN_size + ZN_asset + ZN_number + ZN_money)

Factor 2

(ZN_kind + ZN_feature + ZN_aspect + ZN_type + ZN_area + ZN_reason + ZN_characteristic + ZN_way + ZN_thing + ZN_place + ZN_form + ZN_product + ZN_condition + ZN_time + ZN_source + ZN_factor + ZN_activity + ZN_category + ZN_set + ZN_problem + ZN_question + ZN_word + ZN_task + ZN_company + ZN_function)

Factor 3

(ZN_theory + ZJ_social + ZN_process + ZN_culture + ZN_behavior + ZN_environment + ZJ_cultural + ZN_development + ZN_system + ZN_context + ZN_knowledge + ZN_language + ZN_experience + ZN_idea + ZN_skill + ZN_structure + ZJ_basic + ZN_status + ZJ_complex + ZN_race + ZN_approach + ZN_pattern + ZJ_human + ZN_role + ZN_force + ZN_practice + ZN_ability + ZV_maintain + ZN_belief + ZV_help + ZN_power)

Factor 4

(ZV_influence + ZV_affect + ZJ_particular + ZV_change + ZJ_different + ZV_shape + ZJ_certain + ZJ_such + ZN_change + ZN_response + ZV_occur + ZJ_various + ZJ_specific + ZV_reflect + ZV_depend + ZV_determine + ZJ_political + ZV_create + ZJ_individual + ZV_follow + ZJ_economic + ZV_understand + ZN_outcome + ZJ_key + ZJ_important + ZV_see + ZV_improve + ZV_study + ZV_measure + ZJ_major + ZV_cause + ZV_vary + ZJ_same + ZJ_simple)

Factor 5

(ZN_model + ZN_estimate + ZN_result + ZN_analysis + ZN_datum + ZN_effect + ZN_regression + ZN_equation + ZV_estimate + ZN_table + ZN_study + ZN_measure + ZJ_significant + ZN_example + ZN_experiment + ZN_variable + ZN_sample + ZN_evidence + ZJ_similar + ZN_method + ZJ_positive + ZN_test + ZN_relationship + ZN_term + ZN_participant + ZJ_negative + ZN_difference)

Factor 6

(ZJ_large + ZN_world + ZJ_small + ZN_society + ZN_city + ZN_u.s + ZV_live + ZN_economy + ZJ_many + ZN_population + ZN_family + ZJ_american + ZN_community + ZJ_other + ZN_group + ZJ_local + ZN_country + ZN_state + ZJ_most + ZN_part + ZN_animal + ZN_organization + ZN_institution + ZN_body + ZV_grow + ZN_member + ZV_include + ZV_become + ZN_business)

Factor 7

(ZV_suggest + ZJ_empirical + ZJ_consistent + ZV_support + ZJ_recent + ZV_indicate + ZV_show + ZJ_previous + ZV_examine + ZV_focus + ZV_base + ZN_research + ZV_present + ZN_finding + ZV_describe + ZJ_early + ZN_literature + ZV_report + ZJ_genetic + ZV_explain + ZV_come + ZJ_future)

Factor 8

(ZN_woman + ZJ_young + ZJ_old + ZN_child + ZN_man + ZN_age + ZJ_likely + ZV_tend + ZJ_poor + ZN_education + ZV_work + ZN_home + ZJ_more + ZN_individual + ZN_gender + ZN_people + ZN_student + ZN_work + ZV_receive + ZN_school + ZV_find + ZN_life + ZV_compare)

Factor 9

(ZV_learn + ZV_use + ZV_develop + ZJ_new + ZV_allow + ZV_give + ZV_require + ZV_need + ZJ_available + ZV_apply + ZV_get + ZV_share + ZN_information + ZV_make + ZN_use + ZV_produce + ZJ_common + ZV_take) - (ZN_impact)

Factor 10

(ZN_service + ZN_care + ZJ_public + ZV_provide + ZN_program + ZN_health + ZN_access + ZN_support + ZN_government + ZN_benefit + ZJ_good + ZJ_financial + ZN_resource + ZN_policy + ZN_opportunity + ZN_job + ZN_material + ZJ_direct + ZN_good)

Factor 11

(ZJ_next + ZN_day + ZN_year + ZJ_few + ZN_period + ZN_chapter + ZJ_first + ZJ_several + ZN_century + ZV_spend + ZN_section + ZV_consider + ZV_begin + ZV_discuss + ZJ_long)

Appendix 5 – Descriptive statistics, ANOVA, and means plot from 5-cluster solution HCA for all dimensions

Descriptive Statistics

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Factor1	1	241	8.9735	20.00193	1.28844	6.4354	11.5116	-6.73	74.75
	2	76	-1.2026	9.94702	1.14100	-3.4756	1.0704	-6.73	42.24
	3	1206	-4.5940	4.68409	.13488	-4.8587	-4.3294	-6.73	26.05
	4	67	8.2187	16.87817	2.06200	4.1017	12.3356	-6.73	71.08
	5	25	116.8772	36.86263	7.37253	101.6611	132.0933	74.89	191.50
	Total	1615	.0021	18.83032	.46857	-.9170	.9212	-6.73	191.50
Factor2	1	241	6.7039	16.58310	1.06821	4.5997	8.8082	-3.57	102.24
	2	76	2.5525	10.47772	1.20188	.1582	4.9468	-3.57	55.18
	3	1206	-2.3610	3.23654	.09320	-2.5439	-2.1782	-3.57	22.19
	4	67	13.8740	30.14425	3.68271	6.5213	21.2268	-3.57	141.17
	5	25	4.2160	9.19304	1.83861	.4213	8.0107	-3.57	31.49
	Total	1615	-.0017	10.56775	.26296	-.5175	.5141	-3.57	141.17
Factor3	1	241	7.9872	16.68113	1.07453	5.8705	10.1039	-4.68	80.71
	2	76	6.2050	15.07203	1.72888	2.7609	9.6491	-4.68	62.38
	3	1206	-3.1414	3.67326	.10577	-3.3489	-2.9339	-4.68	19.85
	4	67	19.6764	29.18361	3.56534	12.5580	26.7949	-4.68	154.23
	5	25	2.9012	9.67361	1.93472	-1.0919	6.8943	-4.68	22.95
	Total	1615	-.0007	11.53863	.28712	-.5639	.5625	-4.68	154.23
Factor4	1	241	5.8358	11.68729	.75284	4.3527	7.3188	-5.00	56.54
	2	76	.6178	10.59673	1.21553	-1.8037	3.0392	-5.00	60.14
	3	1206	-3.6938	3.38769	.09755	-3.8852	-3.5024	-5.00	14.11
	4	67	41.3424	28.70780	3.50722	34.3400	48.3448	-5.00	134.03
	5	25	9.1772	19.28536	3.85707	1.2166	17.1378	-5.00	55.37
	Total	1615	-.0012	12.68397	.31562	-.6203	.6178	-5.00	134.03
Factor5	1	241	1.9585	9.24266	.59537	.7857	3.1313	-4.19	53.46
	2	76	32.8741	25.40432	2.91408	27.0689	38.6792	-4.19	133.15
	3	1206	-2.7815	3.46403	.09975	-2.9772	-2.5858	-4.19	19.93
	4	67	4.9591	12.10583	1.47896	2.0063	7.9119	-4.19	47.84

	5	25	2.2772	7.54593	1.50919	-.8376	5.3920	-4.19	28.81
	Total	1615	.0032	10.80897	.26897	-.5243	.5308	-4.19	133.15
Factor6	1	241	3.2968	11.08624	.71413	1.8900	4.7036	-4.44	48.62
	2	76	.6878	10.49511	1.20387	-1.7105	3.0860	-4.44	57.67
	3	1206	-2.8580	4.29158	.12358	-3.1004	-2.6155	-4.44	33.14
	4	67	35.8503	35.56082	4.34445	27.1763	44.5243	-4.44	135.95
	5	25	8.0028	13.51738	2.70348	2.4231	13.5825	-4.44	41.47
	Total	1615	.0013	12.39621	.30846	-.6037	.6064	-4.44	135.95
Factor7	1	241	.7499	6.27866	.40444	-.0468	1.5466	-2.95	32.09
	2	76	22.7433	32.10468	3.68266	15.4071	30.0795	-2.95	139.18
	3	1206	-1.9014	3.52541	.10152	-2.1005	-1.7022	-2.95	34.63
	4	67	5.5449	10.65371	1.30156	2.9463	8.1436	-2.95	34.31
	5	25	.6548	5.09791	1.01958	-1.4495	2.7591	-2.95	12.72
	Total	1615	.0025	9.82288	.24443	-.4769	.4819	-2.95	139.18
Factor8	1	241	6.8458	18.12078	1.16726	4.5464	9.1452	-3.44	125.44
	2	76	1.3396	8.30831	.95303	-.5589	3.2381	-3.44	38.86
	3	1206	-2.2922	3.42585	.09865	-2.4857	-2.0987	-3.44	23.55
	4	67	12.5318	17.74840	2.16831	8.2026	16.8610	-3.44	76.26
	5	25	7.0184	13.15663	2.63133	1.5876	12.4492	-3.44	52.44
	Total	1615	.0014	9.72332	.24195	-.4731	.4760	-3.44	125.44
Factor9	1	241	5.6853	13.48560	.86868	3.9741	7.3965	-11.52	74.06
	2	76	2.9320	10.08719	1.15708	.6270	5.2370	-12.60	39.45
	3	1206	-1.8260	2.61384	.07527	-1.9736	-1.6783	-12.52	19.83
	4	67	6.9228	11.51001	1.40617	4.1153	9.7303	-10.97	49.70
	5	25	5.5852	12.37573	2.47515	.4768	10.6936	-10.51	51.85
	Total	1615	-.0035	7.40011	.18414	-.3647	.3577	-12.60	74.06
Factor10	1	241	6.1183	13.76478	.88667	4.3716	7.8649	-2.70	70.69
	2	76	2.5424	15.67691	1.79826	-1.0400	6.1247	-2.70	127.07
	3	1206	-1.7801	2.67455	.07702	-1.9312	-1.6290	-2.70	14.36
	4	67	3.4384	9.66627	1.18092	1.0806	5.7961	-2.70	42.86
	5	25	9.8124	12.31898	2.46380	4.7274	14.8974	-2.70	44.71
	Total	1615	-.0021	7.82661	.19475	-.3841	.3799	-2.70	127.07
Factor11	1	241	4.3775	13.70075	.88254	2.6389	6.1160	-1.95	78.63
	2	76	3.9808	10.52247	1.20701	1.5763	6.3853	-1.95	41.51
	3	1206	-1.2550	2.71037	.07805	-1.4081	-1.1019	-1.95	21.08
	4	67	2.1318	5.59997	.68415	.7659	3.4977	-1.95	15.12
	5	25	.4936	3.82125	.76425	-1.0837	2.0709	-1.95	10.36
	Total	1615	-.0005	6.70956	.16696	-.3280	.3270	-1.95	78.63

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Factor1	Between Groups	391001.697	4	97750.424	868.092	0.000
	Within Groups	181291.956	1610	112.604		
	Total	572293.654	1614			
Factor2	Between Groups	31390.185	4	7847.546	84.877	.000
	Within Groups	148857.000	1610	92.458		
	Total	180247.185	1614			
Factor3	Between Groups	56352.256	4	14088.064	143.071	.000
	Within Groups	158535.688	1610	98.469		
	Total	214887.945	1614			
Factor4	Between Groups	141312.983	4	35328.246	480.585	.000
	Within Groups	118352.510	1610	73.511		
	Total	259665.494	1614			
Factor5	Between Groups	94165.519	4	23541.380	401.482	.000
	Within Groups	94404.284	1610	58.636		
	Total	188569.803	1614			
Factor6	Between Groups	100218.350	4	25054.588	272.925	.000
	Within Groups	147798.515	1610	91.800		
	Total	248016.865	1614			
Factor7	Between Groups	45877.664	4	11469.416	168.091	.000
	Within Groups	109855.652	1610	68.233		
	Total	155733.316	1614			
Factor8	Between Groups	29520.961	4	7380.240	96.547	.000
	Within Groups	123071.266	1610	76.442		
	Total	152592.227	1614			
Factor9	Between Groups	16454.947	4	4113.737	92.077	.000
	Within Groups	71930.404	1610	44.677		
	Total	88385.351	1614			
Factor10	Between Groups	16533.221	4	4133.305	80.825	.000
	Within Groups	82333.690	1610	51.139		
	Total	98866.910	1614			
Factor11	Between Groups	8032.476	4	2008.119	50.027	.000

	Within Groups	64626.930	1610	40.141		
	Total	72659.406	1614			

Means Plot

Appendix 6 – Dendrograms and Agglomeration Schedule Distance Graphs from HCAs per factor

Factor 1

Factor 2

Agglomeration Schedule - Factor 2
Distance between coefficients
120,000
100,000
80,000
60,000
40,000
20,000
0
1
2
3
4
5
6
7
8
9
10
11
12
Number of clusters

Factor 3

Agglomeration Schedule - Factor 3
Distance between coefficients
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Number of clusters

Factor 4

Agglomeration Schedule - Factor 4
Distance between coefficients
200,000
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
1
2
3
4
5
6
7
8
9
10
11
12
Number of clusters

Factor 5

Factor 6

Agglomeration Schedule - Factor 6
Distance between coefficients
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
1
2
3
4
5
6
7
8
9
10
11
12
Number of clusters

Factor 7

Agglomeration Schedule - Factor 7
Distance between coefficients
100,000
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
1
2
3
4
5
6
7
8
9
10
11
12
Number of clusters

Factor 8

Agglomeration Schedule - Factor 8
Distance between coefficients
100,000
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
1
2
3
4
5
6
7
8
9
10
11
12
Number of clusters

Factor 9

Agglomeration Schedule - Factor 9
Distance between coefficients
60,000
50,000
40,000
30,000
20,000
10,000
0
1
2
3
4
5
6
7
8
9
10
11
12
Number of clusters

Factor 10

Agglomeration Schedule - Factor 10
Distance between coefficients
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
1
2
3
4
5
6
7
8
9
10
11
12
Number of clusters

Factor 11

Agglomeration Schedule - Factor 11
Distance between coefficients
50,000
45,000
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
1
2
3
4
5
6
7
8
9
10
11
12
Number of clusters

Appendix 7 – 5-Cluster solution from HCA for all dimensions

Cluster 1 (241 nodes)
J_next, N_year, J_last, N_day, N_time, N_period, N_month, N_week, J_early, J_few, N_decade, V_begin, J_more, N_end, N_hour, V_consider, N_detail, J_past, J_late, N_century, V_discuss, J_several, V_work, V_take, V_give, N_number, V_follow, J_second, N_reason, N_method, N_term, J_less, J_important, N_school, V_occur, J_same, V_focus, V_look, V_come, N_money, N_amount, N_example, V_sell, N_concept, J_various, N_company, N_researcher, J_new, V_understand, N_turn, N_technique, N_place, N_sample, V_need, V_compare, J_old, J_simple, N_resource, N_growth, N_opportunity, J_key, V_find, J_major, V_produce, N_job, N_worker, N_age, N_order, V_develop, N_education, V_define, V_learn, N_interest, V_grow, N_health, N_service, J_public, N_support, N_access, N_government, J_good, V_receive, N_care, J_available, N_program, V_offer, N_good, V_create, J_great, J_financial, N_system, N_sector, J_direct, V_gain, N_benefit, N_food, N_insurance, N_provider, J_medical, N_training, J_limited, V_limit, N_institution, V_improve, V_require, N_control, N_impact, N_measure, N_material, V_make, N_consumer, N_child, V_design, V_run, N_quality, N_source, N_bank, N_product, V_associate, V_influence, J_indirect, V_affect, N_power, N_law, N_business, N_practice, J_easy, N_tool, N_response, N_asset, N_technology, N_development, N_consumption, N_production, J_real, J_natural, N_sale, N_information, J_poor, N_skill, J_physical, N_rule, V_hold, J_own, N_status, V_relate, N_action, V_change, N_woman, J_particular, V_lead, N_choice, N_approach, N_language, N_knowledge, J_similar, N_student, J_common, V_share, N_sense, N_other, N_task, N_belief, N_set, N_participant, V_identify, V_try, V_involve, V_determine, N_principle, N_share, N_force, N_output, J_complex, N_unit, J_young, N_position, J_difficult, V_expect, N_profit, J_likely, J_specific, V_build, N_variety, V_raise, V_call, N_man, N_adult, N_home, N_class, V_tend, N_gender, N_sex, V_experience, V_engage, N_race, N_category, N_size, V_fall, V_mean, V_shape, V_represent, N_region, V_vary, V_play, J_certain, N_variation, N_trait, N_gene, N_stimulus, N_component, V_depend, N_extent, J_basic, J_scientific, J_organizational, N_brain, V_establish, N_demand, J_general, J_traditional, N_stock, J_equal, V_reflect, V_maintain, N_nature, J_biological, N_function, N_circumstance, N_degree, N_shift, N_perception, J_relative, N_supply, N_productivity, J_cognitive, V_invest, V_rise, J_annual, N_revenue, J_environmental
Cluster 2 (76 nodes)
J_first, J_previous, N_chapter, N_section, N_study, V_examine, V_describe, V_see, V_show, V_present, V_study, N_model, V_illustrate, N_research, N_analysis, N_finding, V_include, N_effect, J_recent, N_work, N_column, V_provide, N_result, N_observation, N_experiment, J_main, V_explain, N_evidence, N_datum, N_use, J_empirical, V_use, J_strong, N_idea, J_additional, V_perform, N_estimate, N_hypothesis, V_exist, V_support, N_table, N_theory, V_test, V_obtain, N_equation, J_multiple, N_relationship, V_base, V_estimate, N_regression, V_reveal, N_error, V_measure, N_statistic, J_statistical, V_generate, V_report, V_ask, V_suggest, N_association, V_demonstrate, N_literature, J_consistent, N_conclusion, V_indicate, N_coefficient, N_correlation, N_interaction, V_conduct, V_apply, N_parameter, J_linear, V_observe, J_positive, J_significant, J_negative
Cluster 3 (1206 nodes)
N_beginning, J_twentieth, V_let, N_half, N_step, J_short, N_night, N_history, N_following, N_dollar, J_entire, N_billion, N_stage, N_topic, V_spend, V_introduce, N_discussion, N_book, V_explore, V_contain, J_nineteenth, N_run, N_generation, V_end, J_long, J_much, N_distance, N_interval, N_lot, N_middle, V_decline, V_emerge, N_career, N_implication, N_length, N_list, N_case, N_lag, J_european, N_duration, N_text, N_tradition, N_point, N_page, N_minute, J_following, N_average, V_move, J_little, N_line, V_turn, V_review, N_exception, N_series, J_critical, V_happen, V_go, V_name, V_note, V_conclude, J_current, N_possibility, J_alternative, N_quarter, V_cover, N_recall, V_appear, V_fix,

V_mention, J_final, N_advantage, V_remain, N_thousand, V_analyze, V_read, V_leave, V_recall, V_summarize, N_phase, J_theoretical, N_million, J_least, N_name, V_seem, N_person, V_date, N_marriage, J_educational, J_foreign, J_private, V_implement, N_exchange, J_raw, N_loss, N_employment, V_aim, N_treatment, J_free, N_trade, J_further, J_federal, V_get, N_distribution, N_concern, J_national, N_security, N_management, N_crisis, V_set, N_contact, N_internet, V_promote, N_connection, J_necessary, N_comparison, J_primary, N_computer, N_customer, N_link, N_artifact, V_ensure, V_lend, N_object, J_potential, N_television, N_opinion, V_buy, N_lack, N_housing, J_mental, N_search, V_maximize, V_bring, J_monetary, V_exploit, J_long-term, V_trade, V_purchase, N_delivery, V_realize, J_international, J_scarce, N_challenge, N_spending, V_lose, V_miss, V_seek, N_space, N_display, N_allocation, N_debate, V_allocate, N_relation, N_awareness, N_industry, J_military, J_domestic, N_agency, N_project, N_manufacturing, N_official, N_professional, N_maker, J_valuable, N_offer, J_bad, N_regulation, N_safety, V_impose, N_intervention, N_insight, N_strategy, N_employee, N_land, J_central, N_subsidy, V_pursue, V_keep, N_disease, N_fit, N_factory, J_global, N_energy, N_deal, V_adopt, N_plan, N_consequence, V_control, N_availability, N_competition, V_acquire, J_rich, N_understanding, V_protect, N_chance, N_news, N_incentive, N_welfare, N_answer, N_water, V_treat, N_framework, J_useful, N_goal, N_indicator, J_sufficient, V_feel, V_argue, N_explanation, N_basis, J_detailed, V_achieve, N_feedback, V_think, N_protection, N_memory, N_grade, N_test, N_description, N_friend, N_record, V_believe, V_know, V_choose, N_effort, N_means, N_fund, J_online, N_view, N_something, N_processing, N_ancestor, N_word, V_start, N_assumption, N_plant, V_allow, N_option, N_logic, N_drug, N_descent, N_field, N_alternative, V_distinguish, N_argument, N_element, N_author, N_theme, N_criterion, N_nation, J_married, N_science, N_account, N_alcohol, V_want, N_offspring, V_address, J_widespread, N_rise, J_efficient, J_extensive, J_hard, N_credit, N_help, N_look, N_cell, V_solve, V_decide, N_birth, V_grant, N_attention, J_special, N_note, N_course, N_difficulty, N_perspective, J_successful, J_considerable, V_speak, J_enough, V_distribute, N_responsibility, V_process, J_possible, V_collect, J_relevant, N_right, V_gather, N_weight, J_visual, V_regard, N_payment, N_procedure, N_flow, N_piece, J_sensory, N_judgment, J_sure, N_contribution, N_inference, V_open, J_full, N_loan, V_extract, V_add, V_encode, N_specie, V_enter, N_distinction, J_clear, V_integrate, N_immigrant, N_communication, J_accurate, N_living, J_standard, N_claim, V_calculate, V_communicate, N_discovery, V_form, N_debt, N_progress, N_one, N_prediction, N_adjustment, N_purchase, N_design, N_statement, N_hand, N_car, N_entry, N_location, J_genetic, N_innovation, N_introduction, N_college, J_male, V_attend, N_parent, V_teach, N_male, J_single, N_neighborhood, N_boy, J_black, N_ethnicity, N_girl, N_teacher, V_die, J_white, N_mother, V_marry, J_middle, V_differ, J_sexual, N_team, N_death, V_bear, V_kill, N_female, V_stay, V_return, V_survive, V_say, V_favor, J_formal, N_mortgage, N_inequality, N_gap, N_identity, V_concentrate, V_rely, J_rural, J_wealthy, N_human, N_office, N_poverty, V_replace, V_own, J_gay, N_equality, J_everyday, J_racial, V_drop, J_secondary, J_mean, N_hospital, N_compensation, N_message, N_site, N_instruction, N_reward, J_daily, V_reach, N_range, N_university, N_score, V_train, J_future, N_cycle, J_female, N_scientist, N_colleague, N_pay, V_save, V_participate, V_possess, J_prior, J_personal, J_urban, N_abuse, V_respond, V_carry, N_father, N_participation, N_minority, N_animal, N_wife, J_indigenous, N_color, J_healthy, V_send, V_watch, N_infant, V_care, N_household, N_attitude, J_modern, N_fact, J_native, N_article, N_review, N_paper, N_notion, N_interpretation, N_respondent, N_phenomenon, N_direction, J_interesting, N_survey, J_psychological, N_body, N_diversity, N_influence, N_investigation, N_expectation, N_play, J_present, N_mechanism, J_existing, N_presence, V_attempt, V_predict, N_variance, N_origin, N_attempt, N_past, N_picture, N_childhood, N_trial, N_trend, N_economics, V_concern, J_able, V_confirm, V_contribute, N_date, J_individual, N_medium, N_advance, N_report, N_figure, V_draw, J_related, V_emphasize, N_panel, V_highlight, V_interpret, N_importance, V_generalize, N_era, N_light, N_graph, N_sign, V_account, J_robust, V_extend, N_patient, J_ethnographic, J_anthropological, N_subject, J_experimental, J_subsequent, J_clinical, N_setting, V_publish, N_resident, N_today, N_scale, J_asian, J_african, N_town, N_percentage, N_fraction, N_proportion, N_whole, J_big, N_tree, N_authority, V_operate, N_leader, J_western, N_corporation, N_anthropologist, N_quantity, J_dominant, J_wild, J_indian, V_organize, N_primate, N_street, N_crop, V_locate, V_eat, N_temperature, N_shape, V_hunt, V_continue, J_japanese, N_citizen, J_steady, N_mind, N_face, J_professional, J_regional, J_nuclear, N_pair, J_internal, N_standard, N_center, J_effective, J_due, N_board, N_owner, J_apparent, V_agree, J_familiar, J_popular, N_rest, N_farmer, J_aware, J_emotional, N_network, N_union, J_industrial, J_essential, J_corporate, N_story, N_executive,

J_interested, J_dependent, J_active, N_manager, J_contemporary, N_density, J_outside, J_expensive, N_organism, J_ethnic, N_house, N_village, N_religion, N_band, N_host, N_room, J_close, J_whole, N_island, N_segment, N_mean, J_overall, N_majority, N_membership, N_item, N_specification, N_game, N_portion, N_target, V_belong, N_economist, J_religious, J_distinct, J_diverse, V_join, N_scholar, N_side, J_random, V_yield, N_magnitude, N_feeling, N_emotion, J_classic, V_employ, V_construct, V_correlate, J_independent, V_propose, V_match, J_causal, V_moderate, V_imply, V_suppose, J_historical, V_fit, J_comparative, J_structural, J_substantial, J_observed, N_laboratory, N_significance, V_derive, N_moderator, N_likelihood, N_mood, V_pass, N_square, J_ordinary, N_show, N_intelligence, N_reaction, V_assign, N_stereotype, V_tell, V_refer, V_complete, J_explanatory, V_capture, V_list, J_precise, V_bias, V_restrict, V_select, V_face, V_compute, N_sensitivity, N_reduction, N_summary, N_present, N_background, V_express, N_partner, N_decline, N_bias, V_display, J_appropriate, N_similarity, J_absolute, V_state, N_meaning, J_net, N_strength, N_data, V_drive, J_null, J_cross-sectional, N_favor, V_exclude, N_collection, J_only, J_continuous, J_constant, V_assume, V_accumulate, J_third, J_dynamic, V_specify, N_anthropology, N_version, N_equilibrium, V_accord, J_behavioral, J_actual, J_estimated, V_attribute, V_denote, V_cause, N_cause, N_probability, J_original, N_testing, N_heterogeneity, J_differential, V_detect, N_climate, N_preference, N_stress, J_true, N_shock, N_dimension, N_frequency, N_success, J_collective, N_damage, N_efficiency, N_learning, N_illness, N_need, N_selection, N_path, V_meet, N_matter, N_ratio, N_depression, N_combination, N_player, N_image, N_index, V_adapt, N_party, J_physiological, V_evolve, J_correct, J_adaptive, N_violence, V_desire, N_pressure, N_nothing, N_barrier, J_slow, N_disorder, N_evolution, N_purpose, N_species, V_remember, N_candidate, J_crucial, N_campaign, V_recognize, J_legal, N_movement, N_dynamics, N_consideration, J_technological, N_everyone, J_rapid, N_transaction, N_recession, J_dramatic, N_reform, V_result, J_sensitive, N_style, N_norm, J_evolutionary, N_hierarchy, N_adaptation, V_underlie, J_criminal, N_justice, V_perceive, J_powerful, N_police, J_false, J_broad, N_balance, N_motor, J_stable, J_institutional, V_lack, J_external, V_reject, N_wealth, J_english, V_enhance, N_application, N_base, N_speaker, V_put, V_interact, V_accept, J_subjective, J_temporal, V_model, N_sound, N_speech, V_expand, V_encourage, V_facilitate, V_prevent, V_avoid, N_migration, J_unique, J_immediate, V_manage, J_prominent, N_formation, N_being, N_banking, J_opposite, J_typical, V_exhibit, J_aggressive, J_regular, N_leadership, V_repeat, J_ongoing, N_classification, N_capacity, J_universal, N_mobility, N_psychology, N_existence, J_fundamental, J_animal, N_tie, N_cognition, N_construction, N_actor, N_isolation, N_complexity, J_distinctive, J_right, J_obvious, J_demographic, J_practical, V_pose, V_arise, N_utility, V_serve, V_answer, N_sentence, J_functional, V_place, V_divide, V_link, V_classify, J_neural, J_geographic, N_surface, J_separate, N_solution, V_write, V_combine, J_french, J_serious, J_extreme, J_vast, N_forest, V_satisfy, J_open, V_surround, J_wrong, J_illegal, V_deal, N_boundary, J_agricultural, V_monitor, N_discount, V_behave, V_encounter, V_eliminate, V_act, N_unemployment, V_lower, J_risky, V_hire, N_inflation, N_investor, N_equity, V_decrease, N_sum, N_premium, N_input, J_top, N_margin, J_marginal, N_uncertainty, J_competitive, N_liability, V_exceed, N_portfolio, N_earnings, J_minimum, N_gain, N_mortality, N_crime, N_division, N_dividend, V_finance, J_willing, V_borrow, N_employer, N_expenditure, N_punishment, J_mass, N_bill, N_exposure, N_conflict, N_fee, N_cash, N_property, J_productive, V_expose, V_supply, N_emission, J_upper, N_decrease, J_huge, N_burden, V_range, V_constitute, V_charge, V_count, N_phone, N_limit, J_sharp, N_elasticity, J_aggregate, J_optimal, N_transportation, V_minimize, N_oil, N_heart, V_equal, V_adjust, N_failure, N_volatility, N_divorce, N_incarceration, V_compete, N_fall, N_rating, N_concentration, N_operation, J_academic, J_adult, J_advanced, J_ancient, J_civil, J_clean, J_cold, J_commercial, J_complete, J_deep, J_eastern, J_ethical, J_executive, J_exogenous, J_explicit, J_fast, J_former, J_fossil, J_fourth, J_german, J_identical, J_implicit, J_impossible, J_informal, J_initial, J_light, J_maximum, J_moral, J_normal, J_northern, J_prime, J_rational, J_reasonable, J_red, J_regulatory, J_reproductive, J_responsible, J_retail, J_sapiens, J_serial, J_severe, J_short-term, J_southern, J_strategic, J_superior, J_surprising, J_systematic, J_tropical, J_unlikely, J_vertical, J_very, J_violent, J_weak, J_wide, J_worth, N_absence, N_achievement, N_act, N_activation, N_advertising, N_agent, N_air, N_alliance, N_anxiety, N_anything, N_ape, N_arm, N_art, N_assessment, N_attack, N_attraction, N_attribute, N_automobile, N_baby, N_bird, N_bit, N_black, N_blood, N_bond, N_bone, N_border, N_bottom, N_branch, N_brand, N_brother, N_budget, N_building, N_buyer, N_carbon, N_card, N_check, N_chromosome, N_civilization, N_clothing, N_confidence, N_constraint, N_contrast, N_cortex, N_couple, N_court, N_curve, N_daughter, N_deficit, N_definition, N_department,

N_deposit, N_desire, N_deviation, N_diet, N_director, N_disadvantage, N_disaster, N_discipline, N_discrimination, N_disparity, N_doctor, N_domain, N_dominance, N_driver, N_earth, N_edge, N_election, N_emphasis, N_estimation, N_evaluation, N_export, N_expression, N_eye, N_fieldwork, N_focus, N_foot, N_freedom, N_fruit, N_fuel, N_future, N_gold, N_gorilla, N_ground, N_habitat, N_head, N_height, N_hundred, N_hunting, N_husband, N_ice, N_identification, N_import, N_improvement, N_instrument, N_interview, N_judge, N_latter, N_letter, N_limitation, N_lobe, N_manipulation, N_measurement, N_meat, N_meeting, N_mile, N_monkey, N_movie, N_music, N_nonhuman, N_officer, N_orientation, N_permission, N_personality, N_personnel, N_politics, N_potential, N_president, N_prison, N_producer, N_professor, N_public, N_radio, N_reader, N_reference, N_relative, N_representation, N_requirement, N_respect, N_restriction, N_retrieval, N_ritual, N_river, N_road, N_root, N_scenario, N_scene, N_scope, N_season, N_seed, N_segregation, N_seller, N_sequence, N_session, N_sheet, N_signal, N_skin, N_slope, N_someone, N_son, N_sport, N_spread, N_stone, N_store, N_substitution, N_survival, N_symbol, N_tendency, N_therapy, N_thought, N_threat, N_top, N_trading, N_tv, N_variability, N_vehicle, N_victim, N_voice, N_war, N_white, V_afford, V_arrive, V_attract, V_break, V_challenge, V_characterize, V_close, V_commit, V_connect, V_consist, V_consume, V_cut, V_depict, V_direct, V_enable, V_evaluate, V_fail, V_feed, V_fight, V_fill, V_forage, V_govern, V_guide, V_hear, V_ignore, V_imagine, V_issue, V_lie, V_like, V_motivate, V_occupy, V_prefer, V_prepare, V_press, V_prove, V_reproduce, V_rule, V_shift, V_stand, V_store, V_suffer, V_travel, V_wait, V_wear, V_win, V_investigate, V_assess
Cluster 4 (67 nodes)
N_issue, N_factor, N_problem, N_way, N_policy, J_different, N_people, N_question, N_country, N_life, V_live, N_process, J_most, N_feature, N_decision, J_other, N_kind, N_characteristic, N_group, N_aspect, V_become, N_individual, J_many, N_part, N_thing, N_condition, N_difference, J_such, J_economic, J_social, V_help, N_family, J_political, N_state, N_performance, J_human, N_outcome, N_culture, N_economy, J_local, N_change, N_role, N_type, N_population, N_community, N_environment, N_form, N_ability, N_experience, J_large, N_pattern, N_member, N_behavior, J_cultural, N_world, N_context, N_situation, N_structure, N_variable, N_organization, N_society, J_small, J_american, N_event, N_activity, N_area, N_city
Cluster 5 (25 nodes)
N_rate, N_income, J_low, J_average, N_firm, N_percent, V_pay, V_earn, N_cost, N_investment, V_reduce, N_market, N_labor, V_increase, N_tax, N_wage, N_price, N_risk, N_capital, N_value, J_high, N_level, N_increase, N_return, J_total

Nodes are preceded by their part-of-speech. J=Adjective; N=Noun; V=Verb

Appendix 8 – Unique collocations per dimension

Collocational Dimension 1: Finance and economics (1,771 collocations)
J_able - V_pay, J_absolute - J_relative, J_absolute - N_value, J_additional - N_cost, J_additional - V_raise, J_aggregate - N_level, J_agricultural - N_labor, J_agricultural - N_production, J_annual - J_average, J_annual - N_growth, J_annual - N_income, J_annual - N_increase, J_annual - N_percent, J_annual - N_rate, J_available - N_amount, J_average - J_high, J_average - J_less, J_average - J_low, J_average - N_amount, J_average - N_firm, J_average - N_growth, J_average - N_income, J_average - N_increase, J_average - N_level, J_average - N_number, J_average - N_price, J_average - N_rate, J_average - N_return, J_average - N_size, J_average - N_value, J_average - N_wage, J_average - N_worker, J_basic - N_level, J_black - N_percent, J_certain - N_amount, J_close - V_pay, J_cognitive - V_associate, J_common - N_share, J_competitive - N_firm, J_competitive - N_market, J_considerable - N_amount, J_constant - N_rate, J_constant - N_return, J_corporate - N_profit, J_corporate - N_rate, J_corporate - N_tax, J_cultural - N_value, J_current - N_price, J_current - N_rate, J_daily - J_average, J_different - N_amount, J_different - N_firm, J_different - N_level, J_different - N_number, J_different - N_rate, J_different - N_size, J_different - N_value, J_different - V_associate, J_different - V_lead, J_direct - N_investment, J_domestic - N_market, J_domestic - N_production, J_dramatic - N_increase, J_due - J_low, J_due - N_increase, J_economic - N_cost, J_economic - N_growth, J_economic - N_level, J_economic - N_profit, J_economic - N_value, J_economic - V_increase, J_economic - V_lead, J_educational - N_level, J_effective - N_rate, J_effective - N_tax, J_effective - V_reduce, J_efficient - J_less, J_efficient - N_market, J_enough - N_money, J_equal - N_cost, J_equal - N_number, J_equal - N_percent, J_equal - N_rate, J_equal - N_share, J_equal - N_value, J_federal - N_level, J_federal - N_tax, J_female - N_worker, J_few - N_firm, J_few - N_worker, J_financial - N_asset, J_financial - N_capital, J_financial - N_cost, J_financial - N_firm, J_financial - N_investment, J_financial - N_market, J_financial - N_risk, J_financial - V_increase, J_foreign - N_investment, J_foreign - N_market, J_free - N_market, J_future - N_price, J_future - N_rate, J_future - N_return, J_future - N_value, J_future - V_expect, J_global - N_market, J_good - N_firm, J_good - N_interest, J_good - N_investment, J_good - N_price, J_great - N_amount, J_great - N_cost, J_great - N_demand, J_great - N_interest, J_great - N_number, J_great - N_rate, J_great - N_risk, J_great - N_value, J_great - V_associate, J_great - V_lead, J_high - J_average, J_high - J_low, J_high - N_capital, J_high - N_cost, J_high - N_demand, J_high - N_firm, J_high - N_growth, J_high - N_income, J_high - N_interest, J_high - N_investment, J_high - N_level, J_high - N_market, J_high - N_number, J_high - N_percent, J_high - N_price, J_high - N_production, J_high - N_profit, J_high - N_rate, J_high - N_return, J_high - N_risk, J_high - N_tax, J_high - N_value, J_high - N_wage, J_high - N_worker, J_high - V_associate, J_high - V_expect, J_high - V_lead, J_high - V_pay, J_high - V_reduce, J_huge - N_amount, J_human - N_capital, J_human - N_growth, J_human - N_labor, J_individual - N_level, J_industrial - N_production, J_international - N_market, J_large - N_amount, J_large - N_cost, J_large - N_firm, J_large - N_income, J_large - N_increase, J_large - N_investment, J_large - N_market, J_large - N_number, J_large - N_price, J_large - N_share, J_large - N_size, J_large - N_value, J_large - V_lead, J_less - J_average, J_less - N_cost, J_less - N_income, J_less - N_percent, J_less - N_price, J_less - N_risk, J_less - N_worker, J_less - V_pay, J_limited - N_amount, J_limited - N_number, J_little - N_money, J_local - N_level, J_local - N_market, J_long-term - N_rate, J_long - J_low, J_low - J_average, J_low - J_high, J_low - N_cost, J_low - N_demand, J_low - N_firm, J_low - N_growth, J_low - N_income, J_low - N_interest, J_low - N_investment, J_low - N_labor, J_low - N_level, J_low - N_price, J_low - N_production, J_low - N_rate, J_low - N_return, J_low - N_risk, J_low - N_tax, J_low - N_value, J_low - N_wage, J_low - N_worker, J_low - V_associate, J_low - V_lead, J_low - V_pay, J_many - N_firm, J_many - N_market, J_many - N_worker, J_many - V_lead, J_marginal - N_cost, J_marginal - N_rate, J_mass - N_production, J_mean - N_size, J_mean - N_value, J_minimum - N_wage, J_more - J_high, J_more - J_less, J_more - N_firm, J_more - N_income, J_more - N_money, J_more - N_percent, J_more - N_value, J_more - N_worker, J_more - V_lead, J_more - V_pay, J_most - N_firm, J_much - N_money, J_much - V_pay, J_national - N_income, J_national - N_level, J_necessary - N_capital, J_negative -

N_value, J_negative - V_associate, J_negative - V_lead, J_negative - V_reduce, J_net - N_income, J_net - N_value, J_new - N_capital, J_new - N_cost, J_new - N_firm, J_new - N_investment, J_new - N_market, J_new - V_lead, J_new - V_raise, J_old - J_less, J_old - N_worker, J_open - N_market, J_optimal - N_level, J_organizational - N_level, J_other - N_asset, J_other - N_firm, J_other - N_number, J_overall - N_increase, J_overall - N_level, J_overall - N_size, J_overall - V_reduce, J_own - N_interest, J_own - N_money, J_particular - N_interest, J_particular - V_associate, J_physical - N_capital, J_poor - J_low, J_positive - N_value, J_positive - V_associate, J_positive - V_expect, J_positive - V_lead, J_potential - N_risk, J_present - N_value, J_primary - N_interest, J_private - N_firm, J_private - N_market, J_productive - N_worker, J_public - N_interest, J_rapid - N_growth, J_rapid - N_increase, J_real - N_growth, J_real - N_interest, J_real - N_investment, J_real - N_rate, J_real - N_value, J_real - N_wage, J_relative - N_cost, J_relative - N_demand, J_relative - N_increase, J_relative - N_price, J_relative - N_rate, J_relative - N_size, J_relative - N_value, J_risky - N_asset, J_risky - N_investment, J_same - N_amount, J_same - N_firm, J_same - N_level, J_same - N_number, J_same - N_price, J_same - N_rate, J_same - N_size, J_secondary - N_market, J_sharp - N_increase, J_significant - N_amount, J_significant - N_increase, J_significant - N_level, J_significant - N_risk, J_single - N_firm, J_slow - N_growth, J_small - N_amount, J_small - N_firm, J_small - N_increase, J_small - N_number, J_small - N_share, J_small - N_size, J_small - N_value, J_social - J_high, J_social - N_capital, J_social - N_cost, J_social - N_level, J_social - V_increase, J_special - N_interest, J_specific - V_associate, J_substantial - N_amount, J_substantial - N_increase, J_such - J_high, J_such - N_cost, J_such - N_firm, J_such - N_market, J_top - N_income, J_top - N_percent, J_total - N_amount, J_total - N_asset, J_total - N_cost, J_total - N_firm, J_total - N_income, J_total - N_increase, J_total - N_investment, J_total - N_market, J_total - N_number, J_total - N_percent, J_total - N_production, J_total - N_profit, J_total - N_share, J_total - N_size, J_total - N_value, J_total - V_increase, J_traditional - N_value, J_true - N_value, J_upper - J_low, J_valuable - N_asset, J_vast - N_amount, J_white - N_percent, J_willing - V_pay, J_young - N_worker, N_ability - N_firm, N_ability - V_increase, N_ability - V_reduce, N_activity - N_firm, N_activity - N_level, N_activity - V_associate, N_activity - V_increase, N_activity - V_reduce, N_adjustment - N_cost, N_advantage - J_relative, N_advantage - N_cost, N_advantage - N_firm, N_age - J_average, N_age - N_increase, N_age - V_increase, N_amount - J_average, N_amount - J_total, N_amount - N_capital, N_amount - N_increase, N_amount - N_investment, N_amount - N_labor, N_amount - N_money, N_amount - V_increase, N_amount - V_pay, N_amount - V_raise, N_amount - V_reduce, N_analysis - N_level, N_area - J_high, N_asset - J_total, N_asset - N_capital, N_asset - N_firm, N_asset - N_price, N_asset - N_return, N_asset - N_value, N_attention - J_less, N_attention - V_pay, N_average - J_high, N_average - N_percent, N_awareness - V_increase, N_bank - N_asset, N_bank - N_interest, N_bank - N_investment, N_bank - N_money, N_bank - N_rate, N_bank - V_pay, N_base - N_tax, N_behavior - V_increase, N_belief - N_value, N_benefit - J_low, N_benefit - N_cost, N_benefit - N_tax, N_benefit - N_wage, N_benefit - N_worker, N_benefit - V_associate, N_benefit - V_increase, N_benefit - V_pay, N_benefit - V_reduce, N_bias - V_lead, N_bill - V_pay, N_body - N_size, N_brain - V_associate, N_brain - V_increase, N_burden - V_reduce, N_business - N_cost, N_business - N_firm, N_business - N_interest, N_business - N_money, N_business - N_risk, N_capacity - V_increase, N_capital - J_high, N_capital - N_amount, N_capital - N_asset, N_capital - N_cost, N_capital - N_firm, N_capital - N_income, N_capital - N_investment, N_capital - N_labor, N_capital - N_level, N_capital - N_market, N_capital - N_rate, N_capital - N_return, N_capital - V_increase, N_capital - V_raise, N_capital - V_reduce, N_care - N_cost, N_case - N_number, N_cash - V_pay, N_cause - V_lead, N_chance - V_increase, N_chance - V_reduce, N_change - J_relative, N_change - N_demand, N_change - N_income, N_change - N_level, N_change - N_price, N_change - N_rate, N_change - N_size, N_change - N_wage, N_change - V_associate, N_change - V_lead, N_characteristic - V_associate, N_child - N_labor, N_child - N_number, N_child - V_raise, N_city - N_capital, N_class - J_low, N_college - J_high, N_color - N_size, N_company - N_profit, N_company - N_value, N_company - V_increase, N_company - V_pay, N_competition - N_market, N_competition - V_increase, N_concentration - J_high, N_concern - V_raise, N_conclusion - V_lead, N_condition - N_demand, N_condition - N_labor, N_condition - N_market, N_conflict - N_interest, N_construction - N_worker, N_consumer - N_demand, N_consumer - N_income, N_consumer - N_price, N_consumer - V_pay, N_consumption - J_high, N_consumption - N_increase, N_consumption - N_level, N_consumption - N_production, N_consumption - V_increase, N_consumption - V_reduce, N_contribution - J_relative, N_control - N_price, N_control - V_increase, N_control - V_reduce, N_correlation - J_high, N_correlation -

N_size, N_cost - J_high, N_cost - J_less, N_cost - J_low, N_cost - J_relative, N_cost - J_total, N_cost - N_capital, N_cost - N_firm, N_cost - N_increase, N_cost - N_investment, N_cost - N_labor, N_cost - N_price, N_cost - N_production, N_cost - N_risk, N_cost - N_tax, N_cost - V_associate, N_cost - V_increase, N_cost - V_pay, N_cost - V_raise, N_cost - V_reduce, N_country - J_high, N_country - J_low, N_country - N_level, N_country - N_market, N_country - N_number, N_country - N_rate, N_country - N_worker, N_credit - N_tax, N_crime - J_high, N_crime - N_rate, N_culture - N_value, N_day - J_average, N_day - N_number, N_death - N_rate, N_death - V_lead, N_debt - V_pay, N_decision - N_firm, N_decision - N_investment, N_decision - V_lead, N_decline - N_price, N_decline - N_rate, N_decline - V_lead, N_decrease - N_increase, N_degree - J_high, N_demand - J_high, N_demand - J_low, N_demand - J_relative, N_demand - N_increase, N_demand - N_labor, N_demand - N_market, N_demand - N_price, N_demand - N_worker, N_demand - V_increase, N_demand - V_reduce, N_density - J_high, N_development - N_growth, N_development - N_level, N_development - V_lead, N_difference - J_average, N_difference - N_income, N_difference - N_rate, N_difference - N_size, N_difference - N_value, N_difference - N_wage, N_difference - V_lead, N_discount - N_rate, N_disease - N_risk, N_distribution - N_firm, N_distribution - N_income, N_distribution - N_production, N_distribution - N_size, N_diversity - V_increase, N_dividend - V_pay, N_division - N_labor, N_divorce - N_rate, N_dollar - J_total, N_dollar - N_amount, N_dollar - N_value, N_dollar - V_pay, N_earnings - J_high, N_earnings - N_labor, N_economy - N_capital, N_economy - N_growth, N_economy - N_market, N_economy - N_production, N_economy - N_rate, N_education - J_high, N_education - N_income, N_education - N_level, N_effect - J_average, N_effect - N_income, N_effect - N_price, N_effect - N_size, N_effect - N_tax, N_effect - N_wage, N_effect - V_expect, N_effect - V_increase, N_efficiency - N_production, N_efficiency - V_increase, N_efficiency - V_reduce, N_effort - V_increase, N_effort - V_reduce, N_elasticity - N_price, N_emission - V_reduce, N_emotion - V_associate, N_employee - N_firm, N_employee - V_pay, N_employer - N_worker, N_employment - N_income, N_employment - N_wage, N_energy - J_less, N_energy - N_amount, N_equilibrium - N_market, N_equilibrium - N_price, N_equity - N_firm, N_equity - N_return, N_error - N_rate, N_estimate - N_size, N_example - N_firm, N_exchange - N_market, N_exchange - N_rate, N_expenditure - J_total, N_exposure - N_risk, N_face - N_value, N_factor - J_total, N_factor - N_demand, N_factor - N_number, N_factor - N_production, N_factor - N_risk, N_factory - N_worker, N_failure - N_market, N_fall - N_price, N_family - N_income, N_family - N_size, N_feature - V_associate, N_fee - V_pay, N_firm - J_average, N_firm - J_high, N_firm - J_low, N_firm - J_total, N_firm - N_asset, N_firm - N_capital, N_firm - N_cost, N_firm - N_investment, N_firm - N_level, N_firm - N_market, N_firm - N_number, N_firm - N_price, N_firm - N_production, N_firm - N_profit, N_firm - N_share, N_firm - N_size, N_firm - N_value, N_firm - N_worker, N_firm - V_expect, N_firm - V_increase, N_firm - V_pay, N_firm - V_reduce, N_food - N_amount, N_food - N_production, N_force - N_labor, N_force - N_market, N_form - N_money, N_form - V_reduce, N_fraction - J_total, N_frequency - J_high, N_frequency - V_increase, N_function - N_cost, N_function - N_demand, N_function - N_production, N_function - N_value, N_function - V_associate, N_function - V_increase, N_fund - N_investment, N_fund - V_raise, N_gain - N_capital, N_gap - N_wage, N_good - N_cost, N_good - N_labor, N_good - N_market, N_good - N_money, N_good - N_price, N_good - N_production, N_good - N_value, N_government - N_interest, N_government - N_money, N_government - N_tax, N_government - V_increase, N_government - V_pay, N_government - V_reduce, N_group - N_interest, N_group - N_level, N_group - N_number, N_group - N_size, N_growth - J_average, N_growth - J_high, N_growth - J_low, N_growth - N_percent, N_growth - N_rate, N_growth - N_wage, N_growth - V_increase, N_growth - V_lead, N_health - N_risk, N_health - V_associate, N_heart - N_rate, N_hour - N_number, N_hour - N_wage, N_hour - N_worker, N_household - J_average, N_household - N_income, N_household - N_percent, N_housing - N_market, N_housing - N_price, N_idea - N_value, N_impact - N_price, N_impact - V_reduce, N_importance - J_relative, N_incarceration - N_rate, N_incentive - V_reduce, N_income - J_average, N_income - J_high, N_income - J_less, N_income - J_low, N_income - J_total, N_income - N_capital, N_income - N_increase, N_income - N_labor, N_income - N_level, N_income - N_percent, N_income - N_rate, N_income - N_share, N_income - N_tax, N_income - N_wage, N_income - V_increase, N_income - V_reduce, N_increase - J_average, N_increase - J_relative, N_increase - J_total, N_increase - N_amount, N_increase - N_cost, N_increase - N_demand, N_increase - N_growth, N_increase - N_income, N_increase - N_interest, N_increase - N_level, N_increase - N_number, N_increase - N_percent, N_increase - N_price, N_increase - N_production, N_increase - N_profit, N_increase - N_rate, N_increase - N_return, N_increase - N_risk, N_increase - N_share,

N_increase - N_size, N_increase - N_tax, N_increase - N_wage, N_increase - V_associate, N_increase - V_expect, N_increase - V_lead, N_increase - V_reduce, N_index - N_price, N_individual - J_high, N_individual - N_number, N_industry - N_firm, N_industry - N_growth, N_inequality - N_income, N_inequality - N_increase, N_inequality - V_increase, N_inequality - V_reduce, N_inflation - J_high, N_inflation - J_low, N_inflation - N_rate, N_information - N_amount, N_information - N_market, N_input - N_labor, N_input - N_production, N_interest - J_high, N_interest - J_low, N_interest - N_increase, N_interest - N_market, N_interest - N_rate, N_interest - N_tax, N_interest - V_increase, N_interest - V_pay, N_interest - V_raise, N_interest - V_reduce, N_investment - J_high, N_investment - J_low, N_investment - J_total, N_investment - N_amount, N_investment - N_capital, N_investment - N_cost, N_investment - N_firm, N_investment - N_level, N_investment - N_rate, N_investment - N_return, N_investment - N_value, N_investment - V_increase, N_investment - V_raise, N_investment - V_reduce, N_investor - N_return, N_investor - N_risk, N_issue - V_raise, N_item - N_number, N_job - J_high, N_job - J_low, N_job - N_wage, N_job - N_worker, N_job - V_pay, N_labor - J_low, N_labor - N_amount, N_labor - N_capital, N_labor - N_cost, N_labor - N_demand, N_labor - N_income, N_labor - N_market, N_labor - N_production, N_labor - N_tax, N_labor - N_wage, N_labor - N_worker, N_labor - V_increase, N_labor - V_pay, N_labor - V_reduce, N_land - N_amount, N_law - N_tax, N_length - J_average, N_level - J_average, N_level - J_high, N_level - J_low, N_level - N_capital, N_level - N_firm, N_level - N_income, N_level - N_increase, N_level - N_investment, N_level - N_percent, N_level - N_price, N_level - N_production, N_level - N_rate, N_level - N_risk, N_level - N_wage, N_level - V_associate, N_level - V_expect, N_level - V_increase, N_liability - N_asset, N_life - N_value, N_life - V_lead, N_likelihood - V_increase, N_likelihood - V_reduce, N_limit - N_amount, N_loan - N_interest, N_loan - N_rate, N_loan - V_pay, N_loss - N_capital, N_loss - V_lead, N_lot - N_money, N_man - N_percent, N_management - N_firm, N_manager - N_worker, N_manufacturing - N_firm, N_margin - N_profit, N_market - J_total, N_market - N_capital, N_market - N_demand, N_market - N_firm, N_market - N_interest, N_market - N_labor, N_market - N_money, N_market - N_percent, N_market - N_price, N_market - N_rate, N_market - N_return, N_market - N_risk, N_market - N_share, N_market - N_size, N_market - N_value, N_market - V_increase, N_material - N_labor, N_measure - N_market, N_measure - N_risk, N_model - N_growth, N_money - N_amount, N_money - N_market, N_money - N_tax, N_money - N_value, N_money - V_increase, N_money - V_pay, N_money - V_raise, N_mortality - J_high, N_mortality - N_rate, N_movement - N_price, N_need - V_reduce, N_norm - N_value, N_number - J_average, N_number - J_high, N_number - J_total, N_number - N_firm, N_number - N_increase, N_number - N_share, N_number - N_size, N_number - N_worker, N_number - V_increase, N_number - V_reduce, N_observation - N_number, N_oil - N_price, N_operation - N_firm, N_opportunity - N_cost, N_opportunity - N_investment, N_opportunity - N_market, N_outcome - N_market, N_outcome - V_associate, N_outcome - V_expect, N_outcome - V_lead, N_output - J_total, N_output - N_firm, N_output - N_level, N_output - N_price, N_output - N_production, N_output - V_increase, N_parameter - N_value, N_parent - V_raise, N_participant - N_market, N_participation - N_labor, N_participation - N_rate, N_pattern - N_growth, N_pay - N_worker, N_payment - N_interest, N_payment - N_tax, N_people - J_high, N_people - N_income, N_people - N_money, N_people - N_number, N_people - V_pay, N_percent - J_high, N_percent - J_less, N_percent - J_total, N_percent - N_growth, N_percent - N_income, N_percent - N_increase, N_percent - N_level, N_percent - N_market, N_percent - N_rate, N_percent - N_return, N_percent - N_share, N_percent - N_wage, N_percent - N_worker, N_percent - V_increase, N_percent - V_pay, N_percentage - J_high, N_percentage - J_total, N_percentage - N_income, N_percentage - N_increase, N_percentage - N_rate, N_performance - J_relative, N_performance - N_firm, N_performance - V_associate, N_period - N_growth, N_period - N_number, N_period - N_rate, N_person - J_average, N_phone - N_number, N_point - J_high, N_point - J_low, N_point - N_increase, N_policy - N_interest, N_policy - N_rate, N_policy - N_tax, N_policy - V_lead, N_policy - V_reduce, N_population - J_high, N_population - J_total, N_population - N_growth, N_population - N_increase, N_population - N_percent, N_population - N_rate, N_population - N_share, N_population - N_size, N_population - V_increase, N_portfolio - N_asset, N_position - J_relative, N_possibility - V_raise, N_poverty - J_high, N_poverty - N_level, N_poverty - N_rate, N_poverty - V_reduce, N_power - J_relative, N_power - N_market, N_power - V_increase, N_premium - J_high, N_premium - N_risk, N_price - J_average, N_price - J_high, N_price - J_less, N_price - J_low, N_price - J_relative, N_price - N_asset, N_price - N_cost, N_price - N_demand, N_price - N_firm, N_price - N_increase, N_price - N_level, N_price - N_market, N_price - N_money, N_price - N_profit, N_price - N_rate, N_price -

N_risk, N_price - N_share, N_price - N_value, N_price - N_wage, N_price - V_expect, N_price - V_increase, N_price - V_lead, N_price - V_pay, N_price - V_raise, N_price - V_reduce, N_probability - J_high, N_probability - J_low, N_probability - N_increase, N_probability - V_increase, N_probability - V_reduce, N_problem - V_associate, N_process - N_production, N_product - J_high, N_product - N_demand, N_product - N_firm, N_product - N_labor, N_product - N_market, N_product - N_price, N_product - N_production, N_product - N_value, N_production - J_high, N_production - J_low, N_production - J_total, N_production - N_cost, N_production - N_firm, N_production - N_increase, N_production - N_labor, N_production - N_level, N_production - N_profit, N_production - N_worker, N_production - V_increase, N_production - V_reduce, N_productivity - J_high, N_productivity - J_low, N_productivity - N_growth, N_productivity - N_increase, N_productivity - N_labor, N_productivity - N_level, N_productivity - N_worker, N_productivity - V_increase, N_profit - J_high, N_profit - J_total, N_profit - N_firm, N_profit - N_increase, N_profit - N_price, N_profit - N_production, N_profit - N_share, N_profit - N_wage, N_profit - V_expect, N_profit - V_increase, N_profit - V_reduce, N_program - V_reduce, N_property - N_tax, N_proportion - J_high, N_proportion - J_total, N_punishment - N_capital, N_quality - J_high, N_quality - J_low, N_quantity - J_total, N_quantity - N_price, N_quantity - V_increase, N_question - V_raise, N_range - N_value, N_rate - J_average, N_rate - J_high, N_rate - J_low, N_rate - J_relative, N_rate - N_capital, N_rate - N_growth, N_rate - N_income, N_rate - N_increase, N_rate - N_interest, N_rate - N_investment, N_rate - N_level, N_rate - N_market, N_rate - N_percent, N_rate - N_price, N_rate - N_return, N_rate - N_risk, N_rate - N_tax, N_rate - N_wage, N_rate - V_expect, N_rate - V_increase, N_rate - V_pay, N_rate - V_raise, N_rate - V_reduce, N_rating - J_high, N_ratio - J_average, N_ratio - J_high, N_ratio - J_low, N_ratio - V_increase, N_reason - V_expect, N_reduction - N_size, N_reduction - V_lead, N_region - V_associate, N_relationship - N_size, N_research - V_lead, N_researcher - V_lead, N_resource - N_amount, N_resource - N_firm, N_resource - N_labor, N_response - N_demand, N_response - N_increase, N_response - N_rate, N_response - V_associate, N_return - J_average, N_return - J_high, N_return - J_low, N_return - N_asset, N_return - N_capital, N_return - N_increase, N_return - N_investment, N_return - N_market, N_return - N_percent, N_return - N_rate, N_return - N_risk, N_return - V_expect, N_revenue - J_total, N_revenue - N_tax, N_revenue - V_increase, N_revenue - V_raise, N_rise - N_income, N_rise - N_price, N_rise - N_rate, N_risk - J_high, N_risk - J_less, N_risk - J_low, N_risk - N_cost, N_risk - N_increase, N_risk - N_level, N_risk - N_market, N_risk - N_price, N_risk - N_rate, N_risk - N_return, N_risk - V_associate, N_risk - V_increase, N_risk - V_raise, N_risk - V_reduce, N_sale - J_total, N_sale - N_firm, N_sale - N_market, N_sale - N_percent, N_sale - N_price, N_sale - N_production, N_sale - N_profit, N_sale - N_tax, N_sample - J_total, N_sample - N_size, N_school - J_high, N_score - J_average, N_score - J_high, N_score - J_low, N_sector - N_worker, N_segment - N_market, N_set - N_value, N_shape - N_size, N_share - J_total, N_share - N_firm, N_share - N_income, N_share - N_increase, N_share - N_market, N_share - N_number, N_share - N_percent, N_share - N_price, N_share - N_profit, N_share - V_increase, N_shift - N_demand, N_significance - N_level, N_size - J_average, N_size - J_relative, N_size - J_total, N_size - N_firm, N_size - N_increase, N_size - N_market, N_size - N_number, N_size - V_increase, N_size - V_reduce, N_skill - N_level, N_source - N_income, N_spending - N_increase, N_spending - V_increase, N_standard - J_high, N_standard - V_raise, N_state - J_high, N_state - N_level, N_state - N_tax, N_status - J_high, N_status - J_low, N_stock - J_high, N_stock - N_asset, N_stock - N_capital, N_stock - N_firm, N_stock - N_investment, N_stock - N_market, N_stock - N_percent, N_stock - N_price, N_stock - N_return, N_stock - N_value, N_strategy - N_firm, N_strength - J_relative, N_strength - N_size, N_stress - V_reduce, N_structure - N_firm, N_structure - N_market, N_student - J_high, N_study - N_number, N_study - N_size, N_subsidy - N_tax, N_sum - J_total, N_sum - N_money, N_supply - N_demand, N_supply - N_increase, N_supply - N_labor, N_supply - N_market, N_supply - N_price, N_supply - V_increase, N_supply - V_reduce, N_system - N_production, N_system - N_tax, N_tax - J_high, N_tax - J_low, N_tax - N_cost, N_tax - N_income, N_tax - N_increase, N_tax - N_interest, N_tax - N_labor, N_tax - N_money, N_tax - N_rate, N_tax - V_increase, N_tax - V_pay, N_tax - V_raise, N_tax - V_reduce, N_team - V_lead, N_technology - N_labor, N_technology - N_production, N_technology - V_reduce, N_temperature - J_high, N_term - N_number, N_time - J_high, N_time - J_less, N_time - N_amount, N_time - N_money, N_time - N_number, N_time - N_price, N_time - N_rate, N_time - N_value, N_time - V_increase, N_topic - N_interest, N_trait - V_associate, N_transaction - N_cost, N_transaction - N_market, N_transportation - N_cost, N_turn - V_lead, N_uncertainty - N_risk, N_uncertainty - V_reduce, N_unemployment - J_high, N_unemployment -

J_low, N_unemployment - N_rate, N_unemployment - N_wage, N_union - N_labor, N_union - N_worker, N_unit - J_low, N_unit - N_cost, N_unit - N_increase, N_unit - N_labor, N_unit - N_price, N_unit - N_production, N_use - V_increase, N_utility - V_expect, N_value - J_average, N_value - J_high, N_value - J_low, N_value - J_relative, N_value - J_total, N_value - N_amount, N_value - N_asset, N_value - N_firm, N_value - N_investment, N_value - N_market, N_value - N_money, N_value - N_price, N_value - V_expect, N_value - V_increase, N_variable - N_interest, N_variable - N_level, N_variable - N_number, N_variable - N_value, N_variation - N_amount, N_volatility - N_price, N_wage - J_average, N_wage - J_high, N_wage - J_low, N_wage - N_growth, N_wage - N_income, N_wage - N_increase, N_wage - N_labor, N_wage - N_level, N_wage - N_percent, N_wage - N_price, N_wage - N_profit, N_wage - N_rate, N_wage - N_worker, N_wage - V_increase, N_wage - V_pay, N_wage - V_raise, N_wage - V_reduce, N_way - N_number, N_way - V_reduce, N_wealth - N_income, N_wealth - V_increase, N_weight - J_low, N_weight - N_size, N_woman - N_labor, N_woman - N_number, N_woman - N_percent, N_woman - N_wage, N_work - N_worker, N_work - V_pay, N_worker - J_average, N_worker - J_high, N_worker - J_less, N_worker - J_low, N_worker - N_demand, N_worker - N_firm, N_worker - N_labor, N_worker - N_number, N_worker - N_percent, N_worker - N_production, N_worker - N_wage, N_worker - V_pay, N_world - N_market, N_world - N_percent, N_world - N_rate, N_world - V_lead, N_year - J_less, N_year - N_interest, N_year - N_number, N_year - N_percent, N_year - N_rate, N_year - V_pay, V_account - N_percent, V_achieve - N_level, V_add - N_value, V_adjust - N_price, V_affect - N_demand, V_affect - N_firm, V_affect - N_growth, V_affect - N_income, V_affect - N_market, V_affect - N_price, V_affect - N_rate, V_affect - N_risk, V_affect - N_wage, V_agree - V_pay, V_allow - N_firm, V_associate - J_high, V_associate - J_low, V_associate - N_cost, V_associate - N_increase, V_associate - N_level, V_associate - N_risk, V_associate - V_increase, V_associate - V_reduce, V_base - N_size, V_base - N_value, V_base - V_expect, V_bear - N_cost, V_bear - N_risk, V_bear - V_raise, V_believe - V_lead, V_borrow - N_money, V_buy - N_money, V_buy - N_price, V_calculate - N_rate, V_calculate - N_size, V_calculate - N_value, V_call - N_rate, V_cause - N_increase, V_cause - N_price, V_cause - N_rate, V_change - N_demand, V_change - N_market, V_change - N_price, V_charge - N_price, V_choose - N_firm, V_choose - N_level, V_collect - N_tax, V_compare - N_cost, V_compare - N_percent, V_compare - N_size, V_compete - N_firm, V_compute - J_average, V_compute - N_value, V_conclude - V_lead, V_consider - N_firm, V_constitute - N_percent, V_continue - N_growth, V_continue - V_increase, V_count - N_number, V_cover - N_cost, V_create - N_demand, V_create - N_value, V_decide - N_firm, V_decline - N_percent, V_decline - N_rate, V_decrease - N_increase, V_decrease - N_rate, V_decrease - V_increase, V_depend - N_amount, V_depend - N_demand, V_depend - N_firm, V_depend - N_income, V_depend - N_level, V_depend - N_number, V_depend - N_price, V_depend - N_rate, V_depend - N_size, V_depend - N_value, V_determine - N_amount, V_determine - N_level, V_determine - N_market, V_determine - N_price, V_determine - N_rate, V_determine - N_size, V_determine - N_value, V_develop - N_market, V_differ - N_size, V_divide - J_total, V_divide - N_number, V_drive - N_price, V_drop - N_percent, V_earn - J_average, V_earn - J_high, V_earn - J_less, V_earn - N_amount, V_earn - N_income, V_earn - N_money, V_earn - N_percent, V_earn - N_profit, V_earn - N_rate, V_earn - N_return, V_earn - N_wage, V_earn - N_worker, V_eliminate - V_reduce, V_emerge - N_market, V_employ - N_worker, V_enter - N_firm, V_enter - N_labor, V_enter - N_market, V_equal - N_value, V_estimate - N_cost, V_estimate - N_value, V_exceed - N_cost, V_exceed - N_value, V_expand - N_firm, V_expand - N_market, V_expect - J_high, V_expect - N_firm, V_expect - N_increase, V_expect - N_level, V_expect - N_price, V_expect - N_profit, V_expect - N_rate, V_expect - N_return, V_expect - N_value, V_expect - V_increase, V_expect - V_lead, V_experience - N_growth, V_experience - N_increase, V_expose - N_risk, V_face - N_firm, V_face - N_risk, V_fall - N_income, V_fall - N_level, V_fall - N_percent, V_fall - N_price, V_fall - N_rate, V_finance - N_investment, V_find - N_rate, V_find - N_value, V_find - V_expect, V_fix - N_cost, V_fix - N_level, V_fix - N_price, V_fix - N_rate, V_generate - N_growth, V_generate - N_income, V_generate - N_profit, V_get - N_money, V_get - V_pay, V_give - N_amount, V_give - N_cost, V_give - N_firm, V_give - N_income, V_give - N_level, V_give - N_money, V_give - N_number, V_give - N_price, V_give - N_rate, V_give - N_value, V_give - V_expect, V_grow - N_demand, V_grow - N_number, V_grow - N_percent, V_grow - N_rate, V_grow - N_size, V_grow - V_expect, V_help - V_reduce, V_hire - N_firm, V_hire - N_labor, V_hire - N_worker, V_hold - N_amount, V_hold - N_asset, V_hold - N_share, V_imply - N_increase, V_impose - N_cost, V_impose - N_tax, V_include - N_cost, V_include - N_income, V_include - N_number, V_increase - J_total, V_increase - N_amount, V_increase - N_capital,

V_increase - N_cost, V_increase - N_demand, V_increase - N_firm, V_increase - N_growth, V_increase - N_income, V_increase - N_interest, V_increase - N_investment, V_increase - N_labor, V_increase - N_level, V_increase - N_market, V_increase - N_money, V_increase - N_number, V_increase - N_percent, V_increase - N_price, V_increase - N_production, V_increase - N_profit, V_increase - N_rate, V_increase - N_risk, V_increase - N_share, V_increase - N_size, V_increase - N_tax, V_increase - N_value, V_increase - N_wage, V_increase - V_associate, V_increase - V_expect, V_increase - V_lead, V_increase - V_reduce, V_indicate - J_high, V_indicate - N_number, V_influence - N_size, V_invest - N_amount, V_invest - N_capital, V_invest - N_firm, V_invest - N_money, V_involve - N_cost, V_involve - N_number, V_involve - N_production, V_keep - J_low, V_keep - N_cost, V_keep - N_interest, V_keep - N_price, V_keep - N_wage, V_lead - J_high, V_lead - J_low, V_lead - N_growth, V_lead - N_increase, V_lead - N_price, V_lead - V_expect, V_lead - V_increase, V_leave - N_money, V_limit - N_amount, V_lose - N_money, V_lose - N_worker, V_lower - N_cost, V_lower - N_price, V_lower - N_rate, V_lower - V_raise, V_maintain - J_high, V_make - N_firm, V_make - N_investment, V_make - N_market, V_make - N_money, V_make - N_profit, V_maximize - N_profit, V_maximize - N_value, V_mean - J_high, V_mean - J_less, V_mean - J_low, V_mean - N_increase, V_mean - N_interest, V_measure - J_relative, V_measure - J_total, V_measure - N_amount, V_measure - N_income, V_measure - N_level, V_measure - N_number, V_meet - N_demand, V_minimize - N_cost, V_need - N_amount, V_need - N_firm, V_need - N_money, V_need - V_pay, V_obtain - N_value, V_offer - J_high, V_offer - N_firm, V_offer - N_price, V_operate - N_firm, V_operate - N_level, V_operate - N_market, V_own - N_asset, V_own - N_firm, V_own - N_percent, V_pay - J_high, V_pay - J_less, V_pay - J_low, V_pay - N_amount, V_pay - N_cost, V_pay - N_firm, V_pay - N_interest, V_pay - N_labor, V_pay - N_money, V_pay - N_percent, V_pay - N_price, V_pay - N_rate, V_pay - N_tax, V_pay - N_wage, V_pay - N_worker, V_perceive - N_risk, V_place - N_demand, V_place - N_value, V_pose - N_risk, V_predict - N_increase, V_predict - N_value, V_produce - J_high, V_produce - J_less, V_produce - N_amount, V_produce - N_cost, V_produce - N_firm, V_produce - N_labor, V_produce - N_worker, V_promote - N_growth, V_put - N_money, V_put - N_risk, V_raise - N_amount, V_raise - N_capital, V_raise - N_cost, V_raise - N_interest, V_raise - N_investment, V_raise - N_money, V_raise - N_price, V_raise - N_rate, V_raise - N_risk, V_raise - N_tax, V_raise - N_wage, V_range - N_size, V_reach - N_level, V_receive - J_high, V_receive - J_less, V_receive - N_amount, V_receive - N_firm, V_receive - N_income, V_receive - N_money, V_receive - N_percent, V_receive - N_share, V_receive - N_wage, V_receive - N_worker, V_receive - V_expect, V_reduce - J_high, V_reduce - N_amount, V_reduce - N_capital, V_reduce - N_cost, V_reduce - N_demand, V_reduce - N_firm, V_reduce - N_income, V_reduce - N_increase, V_reduce - N_interest, V_reduce - N_investment, V_reduce - N_labor, V_reduce - N_number, V_reduce - N_price, V_reduce - N_production, V_reduce - N_profit, V_reduce - N_rate, V_reduce - N_risk, V_reduce - N_size, V_reduce - N_tax, V_reduce - N_wage, V_reduce - V_associate, V_reduce - V_increase, V_reflect - N_cost, V_reflect - N_price, V_relate - N_size, V_remain - N_level, V_report - N_income, V_report - N_number, V_report - N_percent, V_represent - J_total, V_represent - N_amount, V_represent - N_interest, V_represent - N_number, V_represent - N_percent, V_represent - N_value, V_require - N_amount, V_require - N_firm, V_require - N_investment, V_require - N_labor, V_require - N_level, V_require - N_number, V_require - N_rate, V_require - V_pay, V_result - J_high, V_result - J_low, V_result - N_increase, V_rise - N_cost, V_rise - N_demand, V_rise - N_income, V_rise - N_level, V_rise - N_percent, V_rise - N_price, V_rise - N_rate, V_rise - V_expect, V_run - N_risk, V_satisfy - N_demand, V_save - N_money, V_say - N_percent, V_sell - N_asset, V_sell - N_firm, V_sell - N_market, V_sell - N_price, V_serve - N_interest, V_serve - N_market, V_set - J_low, V_set - N_price, V_set - N_rate, V_set - N_wage, V_show - J_average, V_show - J_high, V_show - N_increase, V_show - N_number, V_show - N_rate, V_show - V_increase, V_spend - J_less, V_spend - N_amount, V_spend - N_income, V_spend - N_money, V_supply - N_labor, V_take - N_firm, V_take - N_money, V_take - N_risk, V_take - N_value, V_tend - J_high, V_tend - J_low, V_tend - V_increase, V_use - N_firm, V_use - N_level, V_use - N_money, V_use - N_number, V_use - N_price, V_use - N_rate, V_use - N_value, V_vary - N_amount, V_vary - N_level, V_vary - N_size, V_want - N_money, V_work - J_less, V_work - N_wage, V_work - N_worker
Collocational Dimension 2: Classification (824 collocations)
J_additional - N_factor, J_agricultural - N_product, J_alternative - N_source, J_alternative - N_way, J_available - N_source, J_bad - N_thing, J_basic - N_category, J_basic - N_form, J_basic - N_question, J_basic - N_type, J_biological - N_category, J_biological - N_factor, J_broad -

N_category, J_broad - N_set, J_certain - N_activity, J_certain - N_area, J_certain - N_aspect, J_certain - N_characteristic, J_certain - N_condition, J_certain - N_feature, J_certain - N_kind, J_certain - N_thing, J_certain - N_type, J_certain - N_way, J_cognitive - N_task, J_common - N_factor, J_common - N_feature, J_common - N_form, J_common - N_set, J_complex - N_set, J_complex - N_task, J_criminal - N_activity, J_critical - N_factor, J_critical - N_question, J_crucial - N_question, J_cultural - N_aspect, J_cultural - N_factor, J_daily - N_activity, J_demographic - N_characteristic, J_demographic - N_factor, J_different - N_area, J_different - N_aspect, J_different - N_category, J_different - N_condition, J_different - N_form, J_different - N_function, J_different - N_kind, J_different - N_place, J_different - N_product, J_different - N_question, J_different - N_reason, J_different - N_set, J_different - N_source, J_different - N_task, J_different - N_thing, J_different - N_time, J_different - N_type, J_different - N_way, J_different - N_word, J_difficult - N_problem, J_difficult - N_question, J_difficult - N_task, J_distinct - N_category, J_distinct - N_type, J_distinctive - N_characteristic, J_distinctive - N_feature, J_diverse - N_set, J_domestic - N_product, J_due - N_factor, J_early - N_form, J_easy - N_task, J_easy - N_way, J_economic - N_activity, J_economic - N_condition, J_economic - N_factor, J_economic - N_problem, J_effective - N_way, J_empirical - N_question, J_english - N_word, J_environmental - N_condition, J_environmental - N_factor, J_environmental - N_problem, J_equal - N_thing, J_essential - N_feature, J_experimental - N_condition, J_external - N_factor, J_external - N_source, J_extreme - N_form, J_financial - N_company, J_first - N_condition, J_first - N_place, J_first - N_question, J_first - N_thing, J_first - N_time, J_french - N_word, J_full - N_set, J_functional - N_form, J_fundamental - N_question, J_genetic - N_factor, J_geographic - N_area, J_good - N_company, J_good - N_condition, J_good - N_place, J_good - N_product, J_good - N_reason, J_good - N_thing, J_good - N_time, J_good - N_way, J_great - N_activity, J_great - N_time, J_high - N_area, J_high - N_product, J_high - N_time, J_historical - N_reason, J_human - N_activity, J_human - N_aspect, J_human - N_characteristic, J_human - N_condition, J_human - N_feature, J_illegal - N_activity, J_important - N_area, J_important - N_aspect, J_important - N_characteristic, J_important - N_factor, J_important - N_feature, J_important - N_function, J_important - N_question, J_important - N_reason, J_important - N_source, J_important - N_thing, J_important - N_way, J_individual - N_characteristic, J_interesting - N_question, J_internal - N_factor, J_key - N_area, J_key - N_aspect, J_key - N_characteristic, J_key - N_factor, J_key - N_feature, J_key - N_question, J_large - N_area, J_large - N_company, J_large - N_time, J_least - N_reason, J_less - N_time, J_likely - N_time, J_limited - N_set, J_linear - N_function, J_little - N_reason, J_local - N_area, J_local - N_condition, J_long - N_time, J_long - N_way, J_main - N_reason, J_main - N_source, J_major - N_area, J_major - N_category, J_major - N_factor, J_major - N_problem, J_major - N_reason, J_major - N_source, J_many - N_area, J_many - N_aspect, J_many - N_company, J_many - N_factor, J_many - N_feature, J_many - N_form, J_many - N_kind, J_many - N_place, J_many - N_problem, J_many - N_product, J_many - N_question, J_many - N_reason, J_many - N_thing, J_many - N_time, J_many - N_type, J_many - N_way, J_medical - N_condition, J_medical - N_problem, J_mental - N_problem, J_more - N_company, J_more - N_time, J_most - N_company, J_most - N_time, J_much - N_time, J_multiple - N_factor, J_multiple - N_source, J_natural - N_product, J_necessary - N_condition, J_neural - N_activity, J_new - N_company, J_new - N_form, J_new - N_kind, J_new - N_problem, J_new - N_product, J_new - N_set, J_new - N_source, J_new - N_way, J_obvious - N_question, J_obvious - N_reason, J_only - N_source, J_only - N_thing, J_only - N_way, J_open - N_question, J_other - N_activity, J_other - N_area, J_other - N_aspect, J_other - N_characteristic, J_other - N_company, J_other - N_condition, J_other - N_factor, J_other - N_feature, J_other - N_form, J_other - N_kind, J_other - N_place, J_other - N_problem, J_other - N_product, J_other - N_reason, J_other - N_source, J_other - N_thing, J_other - N_time, J_other - N_type, J_other - N_way, J_other - N_word, J_own - N_set, J_own - N_way, J_particular - N_activity, J_particular - N_area, J_particular - N_aspect, J_particular - N_category, J_particular - N_feature, J_particular - N_form, J_particular - N_kind, J_particular - N_place, J_particular - N_problem, J_particular - N_product, J_particular - N_set, J_particular - N_task, J_particular - N_time, J_particular - N_type, J_particular - N_way, J_physical - N_activity, J_physical - N_characteristic, J_physical - N_feature, J_political - N_activity, J_political - N_factor, J_political - N_form, J_poor - N_condition, J_positive - N_aspect, J_possible - N_reason, J_potential - N_problem, J_potential - N_source, J_practical - N_problem, J_practical - N_reason, J_primary - N_reason, J_primary - N_source, J_psychological - N_factor, J_public - N_place, J_racial - N_category, J_related - N_question, J_relevant - N_aspect, J_relevant - N_factor, J_rich - N_set, J_rich - N_source, J_right - N_place, J_right - N_thing,

J_rural - N_area, J_same - N_area, J_same - N_condition, J_same - N_kind, J_same - N_product, J_same - N_reason, J_same - N_set, J_same - N_thing, J_same - N_time, J_same - N_way, J_same - N_word, J_second - N_reason, J_second - N_set, J_second - N_source, J_second - N_type, J_separate - N_category, J_serious - N_problem, J_several - N_aspect, J_several - N_characteristic, J_several - N_factor, J_several - N_feature, J_several - N_kind, J_several - N_reason, J_several - N_time, J_several - N_way, J_sexual - N_activity, J_sexual - N_characteristic, J_short - N_time, J_similar - N_characteristic, J_similar - N_condition, J_similar - N_feature, J_similar - N_problem, J_similar - N_way, J_simple - N_form, J_simple - N_question, J_simple - N_reason, J_simple - N_task, J_simple - N_thing, J_simple - N_way, J_small - N_area, J_small - N_company, J_small - N_set, J_social - N_aspect, J_social - N_category, J_social - N_condition, J_social - N_factor, J_social - N_form, J_social - N_problem, J_specific - N_area, J_specific - N_aspect, J_specific - N_category, J_specific - N_characteristic, J_specific - N_condition, J_specific - N_feature, J_specific - N_function, J_specific - N_kind, J_specific - N_problem, J_specific - N_question, J_specific - N_set, J_specific - N_task, J_specific - N_type, J_such - N_activity, J_such - N_area, J_such - N_category, J_such - N_characteristic, J_such - N_company, J_such - N_condition, J_such - N_factor, J_such - N_feature, J_such - N_form, J_such - N_place, J_such - N_problem, J_such - N_product, J_such - N_question, J_such - N_task, J_such - N_thing, J_such - N_time, J_such - N_word, J_sufficient - N_condition, J_total - N_factor, J_unique - N_feature, J_urban - N_area, J_useful - N_way, J_various - N_activity, J_various - N_aspect, J_various - N_characteristic, J_various - N_factor, J_various - N_form, J_various - N_kind, J_various - N_reason, J_various - N_source, J_various - N_type, J_various - N_way, J_vast - N_area, J_wrong - N_thing
Collocational Dimension 3: Social and Cultural Development (1,171 collocations)
J_active - N_role, J_actual - N_behavior, J_actual - N_experience, J_actual - N_practice, J_aggressive - N_behavior, J_alternative - N_approach, J_american - N_culture, J_animal - J_human, J_appropriate - N_behavior, J_basic - J_human, J_basic - N_approach, J_basic - N_idea, J_basic - N_knowledge, J_basic - N_process, J_basic - N_skill, J_basic - N_structure, J_basic - N_theory, J_behavioral - N_pattern, J_biological - J_cultural, J_biological - J_human, J_biological - J_social, J_biological - N_process, J_biological - N_race, J_black - N_race, J_broad - N_context, J_causal - N_role, J_central - N_role, J_central - N_system, J_certain - N_behavior, J_clear - N_pattern, J_cognitive - J_complex, J_cognitive - N_ability, J_cognitive - N_development, J_cognitive - N_process, J_cognitive - N_skill, J_cognitive - N_system, J_common - N_approach, J_common - N_belief, J_common - N_culture, J_common - N_language, J_common - N_pattern, J_common - N_practice, J_comparative - N_approach, J_complex - J_social, J_complex - N_behavior, J_complex - N_environment, J_complex - N_knowledge, J_complex - N_pattern, J_complex - N_process, J_complex - N_structure, J_complex - N_system, J_consistent - N_idea, J_consistent - N_pattern, J_consistent - N_theory, J_contemporary - N_culture, J_continuous - N_process, J_criminal - N_behavior, J_criminal - N_system, J_critical - N_development, J_critical - N_role, J_crucial - N_role, J_cultural - J_human, J_cultural - J_social, J_cultural - N_behavior, J_cultural - N_belief, J_cultural - N_context, J_cultural - N_culture, J_cultural - N_development, J_cultural - N_environment, J_cultural - N_force, J_cultural - N_idea, J_cultural - N_knowledge, J_cultural - N_language, J_cultural - N_pattern, J_cultural - N_practice, J_cultural - N_process, J_cultural - N_system, J_current - N_status, J_current - N_system, J_detailed - N_knowledge, J_different - J_cultural, J_different - J_social, J_different - N_approach, J_different - N_behavior, J_different - N_context, J_different - N_culture, J_different - N_environment, J_different - N_language, J_different - N_pattern, J_different - N_process, J_different - N_race, J_different - N_structure, J_different - N_system, J_dynamic - N_process, J_early - J_human, J_early - N_development, J_early - N_experience, J_economic - J_basic, J_economic - J_cultural, J_economic - J_social, J_economic - N_development, J_economic - N_environment, J_economic - N_force, J_economic - N_power, J_economic - N_role, J_economic - N_status, J_economic - N_system, J_economic - N_theory, J_educational - N_system, J_emotional - N_experience, J_english - N_language, J_entire - N_system, J_environmental - J_social, J_environmental - N_context, J_everyday - N_experience, J_everyday - N_practice, J_evolutionary - J_human, J_evolutionary - N_process, J_evolutionary - N_theory, J_explanatory - N_power, J_external - N_environment, J_false - N_belief, J_financial - N_system, J_foreign - N_language, J_formal - N_structure, J_fundamental - J_human, J_general - N_knowledge, J_general - N_pattern, J_general - N_theory, J_genetic - J_human, J_global - N_culture, J_global - N_system, J_good - N_idea, J_good - N_practice, J_great

- J_social, J_great - N_power, J_high - J_social, J_high - N_status, J_high - V_maintain, J_historical - N_context, J_human - J_basic, J_human - J_cultural, J_human - J_social, J_human - N_behavior, J_human - N_culture, J_human - N_development, J_human - N_environment, J_human - N_idea, J_human - N_language, J_human - N_pattern, J_human - N_race, J_human - N_system, J_immediate - N_environment, J_important - J_human, J_important - J_social, J_important - N_role, J_individual - N_ability, J_individual - N_behavior, J_individual - N_experience, J_institutional - N_context, J_internal - N_structure, J_key - N_role, J_large - J_complex, J_large - J_social, J_legal - N_status, J_legal - N_system, J_local - N_culture, J_local - N_environment, J_local - N_knowledge, J_local - N_language, J_low - N_status, J_major - J_cultural, J_major - N_role, J_many - J_cultural, J_many - J_human, J_many - J_social, J_many - N_culture, J_medical - N_practice, J_mental - N_ability, J_mental - N_process, J_military - N_force, J_military - N_power, J_modern - J_human, J_modern - N_system, J_more - J_complex, J_most - J_human, J_native - N_language, J_natural - J_cultural, J_natural - J_human, J_natural - N_behavior, J_natural - N_environment, J_natural - N_process, J_natural - N_theory, J_necessary - N_skill, J_necessary - V_maintain, J_negative - N_experience, J_new - J_social, J_new - N_approach, J_new - N_development, J_new - N_environment, J_new - N_idea, J_new - N_knowledge, J_new - N_practice, J_new - N_theory, J_nuclear - N_power, J_observed - N_pattern, J_ongoing - N_process, J_opposite - N_pattern, J_organizational - N_context, J_organizational - N_practice, J_organizational - N_structure, J_other - J_cultural, J_other - J_social, J_other - N_behavior, J_other - N_culture, J_other - N_language, J_other - N_system, J_own - J_cultural, J_own - N_behavior, J_own - N_belief, J_own - N_culture, J_own - N_experience, J_own - N_knowledge, J_own - N_language, J_particular - J_cultural, J_particular - N_behavior, J_particular - N_context, J_particular - N_culture, J_particular - N_environment, J_particular - N_language, J_past - J_human, J_past - N_experience, J_personal - J_social, J_personal - N_experience, J_physical - J_cultural, J_physical - J_human, J_physical - N_environment, J_political - J_cultural, J_political - J_social, J_political - N_context, J_political - N_culture, J_political - N_environment, J_political - N_force, J_political - N_power, J_political - N_process, J_political - N_role, J_political - N_structure, J_political - N_system, J_poor - V_help, J_popular - N_culture, J_positive - N_experience, J_powerful - N_force, J_previous - N_experience, J_prior - N_experience, J_prior - N_knowledge, J_prominent - N_role, J_psychological - J_social, J_psychological - N_process, J_public - N_system, J_recent - J_human, J_recent - N_development, J_recent - N_experience, J_regular - N_pattern, J_relative - N_power, J_religious - N_belief, J_same - N_pattern, J_scientific - N_approach, J_scientific - N_knowledge, J_scientific - N_theory, J_second - N_approach, J_second - N_language, J_sexual - J_human, J_sexual - N_behavior, J_significant - N_role, J_similar - N_approach, J_similar - N_pattern, J_simple - J_complex, J_simple - N_structure, J_social - J_complex, J_social - J_cultural, J_social - J_human, J_social - N_behavior, J_social - N_context, J_social - N_culture, J_social - N_development, J_social - N_environment, J_social - N_experience, J_social - N_force, J_social - N_pattern, J_social - N_power, J_social - N_process, J_social - N_race, J_social - N_role, J_social - N_skill, J_social - N_status, J_social - N_structure, J_social - N_system, J_social - N_theory, J_social - V_help, J_social - V_maintain, J_specific - J_cultural, J_specific - N_context, J_specific - N_knowledge, J_specific - N_language, J_specific - N_skill, J_stable - V_maintain, J_statistical - N_power, J_strong - N_belief, J_strong - N_force, J_subjective - N_experience, J_such - J_cultural, J_such - J_social, J_such - N_behavior, J_such - N_culture, J_such - N_language, J_such - N_practice, J_such - N_process, J_such - N_system, J_technological - N_development, J_temporal - N_structure, J_theoretical - N_approach, J_theoretical - N_development, J_traditional - N_approach, J_traditional - N_culture, J_traditional - N_practice, J_traditional - N_role, J_traditional - N_system, J_typical - N_pattern, J_unique - N_ability, J_universal - J_human, J_western - N_culture, N_ability - N_language, N_ability - N_power, N_ability - N_skill, N_ability - V_maintain, N_activity - J_human, N_activity - N_pattern, N_actor - J_social, N_adaptation - J_cultural, N_adaptation - N_environment, N_advantage - N_power, N_age - N_race, N_age - N_status, N_animal - J_human, N_anthropologist - J_cultural, N_anthropologist - N_culture, N_anthropology - J_cultural, N_application - N_knowledge, N_approach - J_basic, N_aspect - J_cultural, N_aspect - J_human, N_aspect - J_social, N_aspect - N_behavior, N_aspect - N_culture, N_aspect - N_environment, N_aspect - N_experience, N_aspect - N_language, N_attitude - N_behavior, N_attitude - N_belief, N_authority - N_power, N_background - J_cultural, N_balance - V_maintain, N_banking - N_system, N_base - N_knowledge, N_basis - N_race, N_behavior - J_complex, N_behavior - J_cultural, N_behavior - J_human, N_behavior - J_social, N_behavior - N_belief, N_behavior - N_idea, N_behavior - N_pattern,

N_behavior - N_theory, N_behavior - V_help, N_being - J_human, N_belief - J_cultural, N_belief - N_behavior, N_belief - N_idea, N_belief - N_knowledge, N_belief - N_practice, N_belief - N_system, N_body - J_human, N_body - N_knowledge, N_brain - J_human, N_brain - N_development, N_brain - N_structure, N_capacity - J_human, N_capital - J_human, N_capital - J_social, N_care - N_system, N_category - J_basic, N_category - J_social, N_category - N_race, N_change - J_cultural, N_change - J_human, N_change - J_social, N_change - N_behavior, N_change - N_culture, N_change - N_environment, N_change - N_force, N_change - N_pattern, N_change - N_process, N_change - N_structure, N_characteristic - J_human, N_child - N_ability, N_child - N_behavior, N_child - N_development, N_child - N_skill, N_child - V_help, N_class - J_social, N_class - N_race, N_classification - N_system, N_cognition - J_social, N_communication - J_human, N_communication - N_language, N_communication - N_system, N_complexity - J_social, N_concept - J_basic, N_concept - N_culture, N_concept - N_idea, N_concept - N_race, N_condition - J_human, N_condition - J_social, N_consequence - N_behavior, N_construction - J_social, N_context - J_cultural, N_context - J_social, N_control - J_social, N_control - N_power, N_control - N_process, N_control - V_maintain, N_cost - J_social, N_course - J_human, N_culture - J_human, N_culture - J_social, N_culture - N_environment, N_culture - N_idea, N_culture - N_language, N_culture - N_pattern, N_culture - N_role, N_culture - N_system, N_datum - N_pattern, N_decision - N_process, N_degree - N_power, N_development - J_cultural, N_development - J_human, N_development - J_social, N_development - N_language, N_development - N_process, N_development - N_role, N_development - N_theory, N_difference - J_cultural, N_difference - J_human, N_difference - N_behavior, N_difference - N_status, N_distribution - N_power, N_diversity - J_cultural, N_education - N_skill, N_effect - N_practice, N_environment - J_complex, N_environment - J_cultural, N_environment - J_human, N_environment - J_social, N_environment - N_culture, N_ethnicity - N_race, N_evidence - J_human, N_evidence - N_theory, N_evolution - J_cultural, N_evolution - J_human, N_evolution - N_role, N_evolution - N_theory, N_exchange - J_social, N_exchange - N_system, N_existence - J_human, N_expectation - N_role, N_experience - J_social, N_experience - N_knowledge, N_experience - N_skill, N_factor - J_cultural, N_factor - J_social, N_family - J_social, N_family - N_structure, N_feature - J_human, N_firm - N_ability, N_firm - N_structure, N_force - J_cultural, N_force - J_social, N_form - J_basic, N_form - J_social, N_form - N_behavior, N_formation - N_process, N_function - N_structure, N_gender - N_race, N_gender - N_role, N_gender - N_status, N_gene - J_human, N_gene - N_environment, N_government - N_power, N_government - N_role, N_group - J_cultural, N_group - J_human, N_group - J_social, N_group - N_culture, N_growth - J_human, N_growth - N_development, N_growth - N_pattern, N_health - J_human, N_health - N_status, N_health - N_system, N_hierarchy - J_social, N_hierarchy - N_status, N_history - J_cultural, N_history - J_human, N_history - N_culture, N_human - N_ability, N_hypothesis - N_theory, N_idea - J_basic, N_idea - J_cultural, N_idea - J_human, N_idea - N_behavior, N_idea - N_belief, N_idea - N_culture, N_idea - N_theory, N_identity - J_cultural, N_identity - J_social, N_impact - N_environment, N_importance - N_context, N_incentive - N_system, N_individual - J_social, N_individual - N_ability, N_individual - N_behavior, N_inequality - J_social, N_influence - J_cultural, N_influence - J_social, N_influence - N_power, N_information - N_process, N_information - N_structure, N_information - N_system, N_information - V_help, N_institution - J_cultural, N_institution - J_social, N_institution - N_structure, N_interaction - J_complex, N_interaction - J_human, N_interaction - J_social, N_interaction - N_environment, N_interaction - N_pattern, N_isolation - J_social, N_issue - J_social, N_job - N_skill, N_justice - J_social, N_justice - N_system, N_kind - N_behavior, N_knowledge - J_basic, N_knowledge - J_complex, N_knowledge - J_cultural, N_knowledge - N_belief, N_knowledge - N_experience, N_knowledge - N_skill, N_labor - J_human, N_labor - N_force, N_language - J_cultural, N_language - J_human, N_language - N_ability, N_language - N_culture, N_language - N_development, N_language - N_skill, N_language - N_structure, N_language - N_system, N_leadership - N_role, N_learning - J_social, N_learning - N_skill, N_level - J_basic, N_level - J_social, N_level - N_development, N_level - N_skill, N_life - J_human, N_life - J_social, N_life - N_experience, N_life - N_role, N_man - N_role, N_management - N_practice, N_management - N_system, N_market - N_force, N_market - N_power, N_market - N_structure, N_marriage - N_pattern, N_material - N_culture, N_medium - J_social, N_member - N_culture, N_memory - N_process, N_method - N_knowledge, N_method - N_theory, N_migration - N_pattern, N_mind - N_theory, N_mobility - J_social, N_motor - N_skill, N_movement - J_social, N_nature - J_human, N_need - J_basic, N_need - J_human, N_need - V_help, N_network - J_social, N_network - N_structure, N_norm - J_cultural, N_norm - J_social, N_norm -

N_behavior, N_norm - N_culture, N_order - J_social, N_order - V_maintain, N_organism - N_environment, N_organization - J_complex, N_organization - J_social, N_organization - N_development, N_organization - N_structure, N_origin - J_human, N_other - N_behavior, N_other - V_help, N_outcome - N_behavior, N_outcome - N_process, N_part - J_human, N_part - N_culture, N_part - N_process, N_part - N_system, N_participation - N_force, N_pattern - J_complex, N_pattern - J_cultural, N_pattern - J_human, N_pattern - J_social, N_pattern - N_behavior, N_pattern - N_culture, N_people - J_social, N_people - N_behavior, N_people - N_belief, N_people - N_culture, N_people - N_idea, N_people - V_help, N_person - N_behavior, N_perspective - J_cultural, N_phenomenon - J_social, N_plant - N_power, N_play - N_role, N_police - N_force, N_policy - J_social, N_policy - N_practice, N_policy - V_help, N_population - J_human, N_position - J_social, N_position - N_power, N_position - N_status, N_position - V_maintain, N_power - J_social, N_power - N_ability, N_power - N_status, N_power - N_structure, N_power - N_system, N_power - V_maintain, N_practice - J_cultural, N_practice - N_belief, N_practice - N_theory, N_prediction - N_theory, N_pressure - J_social, N_principle - J_basic, N_problem - J_social, N_problem - N_behavior, N_process - J_basic, N_process - J_complex, N_process - J_cultural, N_process - J_social, N_process - N_development, N_process - N_role, N_process - N_theory, N_product - N_development, N_production - N_process, N_production - N_system, N_program - N_development, N_program - V_help, N_project - N_development, N_psychology - J_social, N_question - J_basic, N_race - J_human, N_race - J_social, N_race - N_status, N_range - J_human, N_relation - J_human, N_relation - J_social, N_relation - N_power, N_relation - V_maintain, N_relationship - J_complex, N_relationship - J_social, N_relationship - N_power, N_relationship - N_status, N_relationship - V_maintain, N_religion - N_culture, N_religion - N_race, N_research - J_basic, N_research - J_social, N_research - N_development, N_research - N_idea, N_research - N_practice, N_research - N_theory, N_resource - J_human, N_resource - J_social, N_resource - N_power, N_response - N_behavior, N_response - N_environment, N_responsibility - J_social, N_result - N_pattern, N_reward - N_system, N_right - J_basic, N_right - J_human, N_role - J_social, N_role - N_culture, N_role - N_development, N_role - N_process, N_role - V_maintain, N_rule - J_cultural, N_school - N_system, N_science - J_social, N_scientist - J_social, N_security - J_social, N_selection - N_process, N_selection - N_role, N_selection - N_theory, N_service - J_social, N_set - J_complex, N_set - N_belief, N_set - N_skill, N_setting - J_cultural, N_setting - J_social, N_sex - N_role, N_sign - N_language, N_situation - J_social, N_skill - J_basic, N_skill - J_social, N_skill - N_ability, N_skill - N_experience, N_skill - N_knowledge, N_skill - N_language, N_society - J_complex, N_society - J_cultural, N_society - J_human, N_society - J_social, N_society - N_culture, N_society - N_role, N_society - N_system, N_sound - N_language, N_source - N_power, N_speaker - N_language, N_specie - J_human, N_species - J_human, N_speech - N_language, N_stage - N_development, N_standard - V_maintain, N_state - N_power, N_status - J_social, N_status - N_power, N_status - N_race, N_step - N_process, N_strategy - N_development, N_structure - J_basic, N_structure - J_complex, N_structure - J_social, N_structure - N_language, N_structure - N_power, N_student - V_help, N_study - J_human, N_study - N_behavior, N_support - J_social, N_support - N_idea, N_system - J_complex, N_system - J_cultural, N_system - J_human, N_system - J_social, N_system - N_belief, N_system - N_culture, N_system - N_language, N_system - N_power, N_system - V_maintain, N_task - J_complex, N_tax - N_system, N_technology - N_development, N_term - J_basic, N_term - N_race, N_term - N_structure, N_test - N_power, N_theory - J_basic, N_theory - J_social, N_theory - N_behavior, N_theory - N_development, N_theory - N_practice, N_theory - N_process, N_theory - V_help, N_tie - J_social, N_tradition - J_cultural, N_training - N_skill, N_trait - J_human, N_trait - N_behavior, N_treatment - N_approach, N_type - J_basic, N_type - N_system, N_understanding - J_human, N_understanding - N_knowledge, N_unit - J_basic, N_use - N_force, N_use - N_language, N_use - N_pattern, N_value - J_cultural, N_value - N_belief, N_value - N_culture, N_value - N_idea, N_variation - J_cultural, N_variation - J_human, N_variation - N_pattern, N_way - N_culture, N_way - N_language, N_way - V_help, N_wealth - N_power, N_welfare - J_social, N_woman - N_experience, N_woman - N_role, N_woman - N_status, N_word - N_language, N_work - N_behavior, N_work - N_environment, N_work - N_experience, N_work - N_force, N_work - N_practice, N_work - N_role, N_work - V_help, N_world - J_social, N_world - N_culture, N_world - N_knowledge, N_world - N_language, N_world - N_system, V_accept - N_idea, V_accord - N_theory, V_account - N_theory, V_accumulate - N_knowledge, V_achieve - N_status, V_achieve - V_help, V_acquire - N_ability, V_acquire - N_knowledge, V_acquire - N_skill, V_adapt - N_ability, V_adapt - N_environment, V_adopt - N_approach, V_affect - J_social, V_affect - N_ability, V_affect - N_behavior,

V_affect - N_context, V_affect - N_environment, V_affect - N_experience, V_affect - N_force, V_affect - N_process, V_affect - N_structure, V_allow - N_approach, V_allow - N_system, V_apply - N_approach, V_apply - N_idea, V_apply - N_knowledge, V_apply - N_theory, V_argue - N_theory, V_assign - N_role, V_assume - N_theory, V_avoid - V_help, V_base - J_social, V_base - N_belief, V_base - N_culture, V_base - N_experience, V_base - N_idea, V_base - N_knowledge, V_base - N_race, V_base - N_status, V_base - N_system, V_base - N_theory, V_become - J_complex, V_begin - N_process, V_build - N_structure, V_build - N_theory, V_build - V_help, V_build - V_maintain, V_call - N_approach, V_call - N_process, V_call - N_structure, V_call - N_system, V_call - N_theory, V_change - N_behavior, V_change - N_culture, V_change - N_environment, V_change - N_pattern, V_change - N_structure, V_come - N_idea, V_come - N_power, V_consider - N_approach, V_contribute - N_development, V_contribute - N_knowledge, V_control - N_ability, V_control - N_power, V_create - J_social, V_create - N_ability, V_create - N_culture, V_create - N_environment, V_create - N_experience, V_create - N_pattern, V_create - N_power, V_create - N_process, V_create - N_structure, V_create - N_system, V_create - V_help, V_create - V_maintain, V_decide - V_help, V_define - N_culture, V_define - N_race, V_depend - N_ability, V_depend - N_context, V_describe - N_approach, V_describe - N_behavior, V_describe - N_experience, V_describe - N_process, V_describe - N_theory, V_design - V_help, V_detect - N_power, V_determine - N_behavior, V_determine - N_role, V_determine - N_structure, V_determine - V_help, V_develop - N_ability, V_develop - N_approach, V_develop - N_culture, V_develop - N_idea, V_develop - N_language, V_develop - N_skill, V_develop - N_system, V_develop - N_theory, V_develop - V_help, V_drive - N_force, V_emerge - N_pattern, V_emphasize - N_approach, V_emphasize - N_role, V_encourage - N_development, V_engage - N_ability, V_engage - N_behavior, V_engage - N_practice, V_engage - N_process, V_enhance - N_ability, V_ensure - V_help, V_establish - N_system, V_establish - V_help, V_establish - V_maintain, V_evolve - J_human, V_examine - J_social, V_examine - N_approach, V_examine - N_pattern, V_examine - N_role, V_exhibit - N_pattern, V_expand - N_knowledge, V_explain - J_human, V_explain - J_social, V_explain - N_behavior, V_explain - N_pattern, V_explain - N_theory, V_explain - V_help, V_explore - N_role, V_express - N_idea, V_facilitate - N_development, V_find - N_pattern, V_fit - N_pattern, V_focus - J_social, V_focus - N_approach, V_focus - N_role, V_focus - N_theory, V_follow - N_approach, V_follow - N_behavior, V_follow - N_pattern, V_follow - N_practice, V_follow - N_process, V_gain - N_experience, V_gain - N_knowledge, V_gain - N_power, V_gain - N_skill, V_generate - N_knowledge, V_get - N_idea, V_give - N_idea, V_give - N_power, V_help - J_social, V_help - N_behavior, V_help - N_theory, V_help - V_maintain, V_hold - N_belief, V_hold - N_power, V_hold - N_theory, V_identify - N_ability, V_identify - N_pattern, V_identify - V_help, V_illustrate - N_approach, V_illustrate - N_idea, V_illustrate - N_process, V_implement - N_practice, V_imply - N_theory, V_improve - N_ability, V_improve - N_practice, V_improve - V_help, V_include - J_human, V_include - J_social, V_include - N_behavior, V_include - N_culture, V_include - N_language, V_include - N_system, V_increase - J_social, V_increase - N_ability, V_increase - N_behavior, V_increase - N_power, V_influence - J_cultural, V_influence - J_social, V_influence - N_ability, V_influence - N_behavior, V_influence - N_culture, V_influence - N_process, V_interact - N_environment, V_interpret - N_behavior, V_introduce - N_idea, V_involve - J_complex, V_involve - J_cultural, V_involve - J_social, V_involve - N_practice, V_involve - N_process, V_involve - N_system, V_join - N_force, V_keep - V_help, V_lack - N_skill, V_lead - N_development, V_learn - N_ability, V_learn - N_behavior, V_learn - N_context, V_learn - N_culture, V_learn - N_environment, V_learn - N_experience, V_learn - N_language, V_learn - N_practice, V_learn - N_process, V_learn - N_skill, V_learn - N_theory, V_limit - N_ability, V_limit - N_power, V_live - N_environment, V_look - N_pattern, V_lose - N_ability, V_maintain - J_social, V_maintain - N_ability, V_maintain - N_power, V_maintain - N_role, V_maintain - N_system, V_maintain - V_help, V_make - N_process, V_make - N_system, V_make - V_help, V_manage - V_help, V_measure - N_ability, V_model - N_approach, V_need - N_skill, V_observe - N_behavior, V_observe - N_pattern, V_occur - N_behavior, V_occur - N_context, V_occur - N_process, V_operate - N_system, V_perceive - J_social, V_perceive - N_ability, V_play - N_role, V_predict - N_behavior, V_predict - N_theory, V_prevent - V_help, V_produce - N_ability, V_produce - N_behavior, V_produce - N_process, V_produce - N_system, V_promote - N_development, V_propose - N_theory, V_protect - N_environment, V_protect - V_help, V_provide - J_social, V_provide - N_environment, V_provide - N_system, V_provide - V_help, V_put - N_practice, V_recognize - N_ability, V_reduce - N_ability, V_reduce - V_help, V_reflect - N_behavior, V_reflect - N_experience, V_reflect - N_pattern, V_reflect - N_process, V_reflect - N_status, V_reflect - N_structure,

V_reject - N_idea, V_relate - N_behavior, V_relate - N_experience, V_repeat - N_process, V_represent - N_pattern, V_require - J_complex, V_require - N_approach, V_require - N_knowledge, V_require - N_skill, V_reveal - N_pattern, V_see - N_behavior, V_see - N_culture, V_see - N_pattern, V_shape - J_cultural, V_shape - J_human, V_shape - J_social, V_shape - N_behavior, V_shape - N_culture, V_shape - N_environment, V_shape - N_experience, V_shape - N_force, V_shape - N_role, V_shape - N_structure, V_share - J_cultural, V_share - N_belief, V_share - N_culture, V_share - N_experience, V_share - N_idea, V_share - N_knowledge, V_show - N_pattern, V_speak - N_language, V_state - N_theory, V_study - J_human, V_study - J_social, V_study - N_approach, V_study - N_behavior, V_study - N_culture, V_study - N_language, V_suggest - N_approach, V_suggest - N_pattern, V_suggest - N_theory, V_support - N_idea, V_support - N_theory, V_survive - N_ability, V_take - N_approach, V_take - N_role, V_teach - N_skill, V_test - N_idea, V_test - N_theory, V_think - N_ability, V_underlie - N_process, V_underlie - N_structure, V_understand - J_basic, V_understand - J_complex, V_understand - J_human, V_understand - J_social, V_understand - N_ability, V_understand - N_approach, V_understand - N_behavior, V_understand - N_context, V_understand - N_culture, V_understand - N_language, V_understand - N_process, V_understand - N_role, V_understand - V_help, V_use - N_approach, V_use - N_knowledge, V_use - N_language, V_use - N_practice, V_use - N_process, V_use - N_system, V_vary - N_context, V_vary - N_culture, V_vary - N_practice, V_work - N_force, V_work - N_system
Collocational Dimension 4: Exerting Influence and Shaping Outcomes (1,227 collocations)
J_able - V_understand, J_actual - N_outcome, J_adaptive - N_response, J_appropriate - N_response, J_bad - N_outcome, J_basic - J_economic, J_basic - V_understand, J_behavioral - N_outcome, J_behavioral - N_response, J_certain - J_such, J_collective - J_individual, J_complex - J_simple, J_complex - V_understand, J_correct - N_response, J_critical - V_understand, J_crucial - V_understand, J_cultural - J_different, J_cultural - J_economic, J_cultural - J_major, J_cultural - J_particular, J_cultural - J_political, J_cultural - J_specific, J_cultural - J_such, J_cultural - N_change, J_cultural - V_influence, J_cultural - V_shape, J_current - J_economic, J_current - V_depend, J_different - J_economic, J_different - J_same, J_different - V_depend, J_different - V_reflect, J_different - V_vary, J_difficult - V_determine, J_difficult - V_measure, J_difficult - V_understand, J_dramatic - N_change, J_due - N_change, J_early - V_see, J_easy - V_measure, J_easy - V_see, J_easy - V_understand, J_economic - J_different, J_economic - J_major, J_economic - J_political, J_economic - J_such, J_economic - N_change, J_economic - N_outcome, J_economic - V_affect, J_economic - V_change, J_economic - V_create, J_economic - V_improve, J_emotional - N_response, J_environmental - N_change, J_environmental - N_response, J_environmental - V_influence, J_essential - V_understand, J_financial - V_improve, J_first - V_follow, J_first - V_occur, J_future - J_important, J_future - N_outcome, J_future - V_depend, J_general - J_specific, J_genetic - N_change, J_global - J_economic, J_global - N_change, J_good - N_outcome, J_good - N_response, J_great - J_economic, J_high - J_such, J_human - J_important, J_human - N_change, J_human - V_shape, J_human - V_study, J_human - V_understand, J_important - J_such, J_important - V_determine, J_important - V_understand, J_individual - V_influence, J_interested - V_study, J_large - J_such, J_large - N_change, J_legal - J_political, J_likely - V_occur, J_local - N_response, J_major - J_economic, J_major - N_change, J_many - J_different, J_many - J_economic, J_many - J_important, J_many - J_same, J_many - J_such, J_many - V_see, J_military - J_political, J_more - J_such, J_more - V_create, J_more - V_see, J_much - V_depend, J_much - V_determine, J_natural - J_such, J_necessary - V_understand, J_negative - N_outcome, J_new - J_such, J_new - V_create, J_next - V_see, J_organizational - N_change, J_other - J_different, J_other - J_economic, J_other - J_same, J_other - J_such, J_other - V_affect, J_other - V_see, J_own - V_create, J_particular - J_specific, J_particular - V_depend, J_personal - J_political, J_physical - J_such, J_physiological - N_response, J_political - J_economic, J_political - N_change, J_political - V_influence, J_political - V_shape, J_poor - N_outcome, J_positive - N_outcome, J_positive - N_response, J_possible - N_outcome, J_public - J_such, J_public - V_influence, J_random - V_follow, J_rapid - N_change, J_relative - N_change, J_relative - V_measure, J_religious - J_political, J_same - J_different, J_sensitive - N_change, J_several - J_different, J_several - J_important, J_several - J_key, J_several - J_major, J_significant - N_change, J_similar - J_different, J_similar - V_occur, J_slow - N_response, J_small - N_change, J_social - J_different, J_social - J_economic, J_social - J_important, J_social - J_political, J_social - J_such, J_social - N_change, J_social - V_affect, J_social - V_create,

J_social - V_influence, J_social - V_shape, J_social - V_study, J_social - V_understand, J_specific - J_particular, J_specific - V_depend, J_structural - N_change, J_such - J_certain, J_such - J_economic, J_such - J_important, J_such - N_change, J_such - N_outcome, J_such - V_occur, J_technological - N_change, J_total - V_measure, J_true - J_same, J_true - V_reflect, J_useful - V_study, J_useful - V_understand, N_ability - J_individual, N_ability - V_affect, N_ability - V_create, N_ability - V_depend, N_ability - V_improve, N_ability - V_influence, N_ability - V_measure, N_ability - V_understand, N_action - J_individual, N_action - J_political, N_action - J_specific, N_action - J_such, N_action - V_cause, N_action - V_influence, N_activity - J_certain, N_activity - J_economic, N_activity - J_particular, N_activity - J_political, N_activity - J_such, N_activity - J_various, N_activity - N_response, N_activity - V_measure, N_activity - V_reflect, N_age - J_different, N_age - J_same, N_amount - J_certain, N_amount - J_different, N_amount - J_same, N_amount - V_depend, N_amount - V_determine, N_amount - V_measure, N_amount - V_vary, N_analysis - J_simple, N_animal - J_such, N_answer - J_simple, N_answer - V_depend, N_anthropologist - V_study, N_approach - J_different, N_approach - V_follow, N_approach - V_study, N_approach - V_understand, N_area - J_certain, N_area - J_different, N_area - J_important, N_area - J_key, N_area - J_major, N_area - J_particular, N_area - J_same, N_area - J_specific, N_area - J_such, N_aspect - J_certain, N_aspect - J_different, N_aspect - J_important, N_aspect - J_key, N_aspect - J_particular, N_aspect - J_specific, N_aspect - J_various, N_assumption - J_key, N_attention - J_particular, N_attitude - V_change, N_barrier - V_create, N_behavior - J_certain, N_behavior - J_different, N_behavior - J_individual, N_behavior - J_particular, N_behavior - J_such, N_behavior - N_change, N_behavior - N_outcome, N_behavior - N_response, N_behavior - V_affect, N_behavior - V_change, N_behavior - V_determine, N_behavior - V_follow, N_behavior - V_influence, N_behavior - V_occur, N_behavior - V_reflect, N_behavior - V_see, N_behavior - V_shape, N_behavior - V_study, N_behavior - V_understand, N_benefit - J_economic, N_brain - J_different, N_business - J_different, N_business - J_political, N_campaign - J_political, N_candidate - J_political, N_care - V_improve, N_case - J_simple, N_case - J_such, N_category - J_different, N_category - J_major, N_category - J_particular, N_category - J_specific, N_category - J_such, N_category - V_create, N_cause - J_major, N_center - J_major, N_challenge - J_major, N_challenge - V_create, N_change - J_economic, N_change - J_major, N_change - J_political, N_change - J_such, N_change - N_response, N_change - V_affect, N_change - V_cause, N_change - V_follow, N_change - V_measure, N_change - V_occur, N_change - V_reflect, N_change - V_see, N_chapter - V_see, N_characteristic - J_certain, N_characteristic - J_important, N_characteristic - J_individual, N_characteristic - J_key, N_characteristic - J_specific, N_characteristic - J_such, N_characteristic - J_various, N_characteristic - V_determine, N_characteristic - V_vary, N_choice - J_individual, N_choice - V_affect, N_choice - V_depend, N_choice - V_influence, N_circumstance - J_certain, N_circumstance - J_particular, N_circumstance - J_specific, N_circumstance - V_depend, N_circumstance - V_vary, N_city - J_major, N_class - J_different, N_climate - N_change, N_climate - V_affect, N_climate - V_change, N_color - J_different, N_combination - J_various, N_community - V_study, N_company - J_such, N_component - J_different, N_component - J_important, N_component - J_key, N_component - J_major, N_concept - J_important, N_concept - J_key, N_concept - V_understand, N_concern - J_major, N_conclusion - V_affect, N_condition - J_certain, N_condition - J_different, N_condition - J_economic, N_condition - J_same, N_condition - J_specific, N_condition - J_such, N_condition - N_change, N_condition - N_response, N_condition - V_change, N_condition - V_create, N_condition - V_improve, N_consequence - J_economic, N_consequence - J_important, N_consideration - J_important, N_context - J_different, N_context - J_particular, N_context - J_political, N_context - J_specific, N_context - V_affect, N_context - V_depend, N_context - V_occur, N_context - V_understand, N_context - V_vary, N_contribution - J_important, N_contribution - J_major, N_cost - J_economic, N_cost - J_such, N_cost - V_reflect, N_country - J_different, N_country - J_economic, N_country - J_particular, N_country - J_political, N_country - J_same, N_country - J_such, N_country - J_various, N_country - V_vary, N_course - V_change, N_crisis - J_economic, N_culture - J_different, N_culture - J_particular, N_culture - J_political, N_culture - J_such, N_culture - N_change, N_culture - V_change, N_culture - V_create, N_culture - V_influence, N_culture - V_see, N_culture - V_shape, N_culture - V_study, N_culture - V_understand, N_culture - V_vary, N_damage - V_cause, N_datum - J_such, N_death - V_cause, N_decision - J_economic, N_decision - J_important, N_decision - J_individual, N_decision - J_political, N_decision - N_outcome, N_decision - V_affect, N_decision - V_depend, N_decision - V_influence, N_degree - J_different, N_degree - V_depend, N_degree - V_reflect, N_degree -

V_vary, N_demand - N_change, N_demand - N_response, N_demand - V_affect, N_demand - V_change, N_demand - V_create, N_demand - V_depend, N_depression - J_major, N_development - J_economic, N_difference - J_economic, N_difference - J_important, N_difference - J_individual, N_difference - J_key, N_difference - J_such, N_difference - N_outcome, N_difference - V_cause, N_difference - V_reflect, N_difference - V_understand, N_dimension - J_key, N_dimension - J_various, N_direction - J_same, N_direction - V_change, N_discussion - V_follow, N_discussion - V_see, N_disease - J_such, N_disease - V_cause, N_disorder - J_specific, N_distinction - J_important, N_distribution - V_affect, N_distribution - V_follow, N_dynamics - V_understand, N_economy - J_political, N_economy - V_affect, N_effect - J_different, N_effect - J_simple, N_effect - J_such, N_effect - N_change, N_effect - N_outcome, N_effect - V_depend, N_effect - V_measure, N_effect - V_occur, N_effect - V_see, N_effect - V_study, N_efficiency - V_improve, N_effort - V_change, N_effort - V_improve, N_element - J_important, N_element - J_key, N_environment - J_different, N_environment - J_economic, N_environment - J_particular, N_environment - J_political, N_environment - N_change, N_environment - N_response, N_environment - V_affect, N_environment - V_change, N_environment - V_create, N_environment - V_shape, N_equilibrium - V_determine, N_estimate - J_different, N_event - J_major, N_event - J_specific, N_event - J_such, N_event - V_affect, N_event - V_cause, N_event - V_follow, N_event - V_influence, N_event - V_occur, N_everyone - J_same, N_evidence - J_such, N_evolution - V_understand, N_example - J_simple, N_example - V_see, N_experience - J_individual, N_experience - V_affect, N_experience - V_create, N_experience - V_reflect, N_experience - V_shape, N_explanation - J_simple, N_extent - V_depend, N_extent - V_determine, N_extent - V_influence, N_extent - V_vary, N_fact - J_important, N_fact - V_reflect, N_factor - J_economic, N_factor - J_important, N_factor - J_key, N_factor - J_major, N_factor - J_political, N_factor - J_such, N_factor - J_various, N_factor - N_change, N_factor - V_affect, N_factor - V_cause, N_factor - V_depend, N_factor - V_determine, N_factor - V_influence, N_factor - V_shape, N_feature - J_certain, N_feature - J_important, N_feature - J_key, N_feature - J_particular, N_feature - J_specific, N_feature - J_such, N_finding - J_important, N_finding - J_key, N_firm - J_different, N_firm - J_same, N_firm - J_such, N_firm - V_affect, N_firm - V_depend, N_food - J_such, N_force - J_economic, N_force - J_political, N_force - N_change, N_force - V_affect, N_force - V_shape, N_form - J_different, N_form - J_particular, N_form - J_political, N_form - J_simple, N_form - J_such, N_form - J_various, N_framework - V_understand, N_frequency - N_change, N_frequency - V_occur, N_function - J_different, N_function - J_important, N_function - J_specific, N_function - N_response, N_function - V_vary, N_gender - J_such, N_gene - J_particular, N_gene - V_determine, N_gene - V_influence, N_goal - J_specific, N_government - J_economic, N_government - J_various, N_government - V_create, N_group - J_certain, N_group - J_different, N_group - J_individual, N_group - J_particular, N_group - J_political, N_group - J_same, N_group - J_such, N_group - J_various, N_group - V_see, N_group - V_vary, N_growth - J_economic, N_growth - V_affect, N_health - N_outcome, N_health - V_affect, N_health - V_improve, N_household - J_individual, N_hypothesis - J_specific, N_illness - V_cause, N_image - V_create, N_impact - J_economic, N_impact - J_major, N_impact - N_change, N_impact - N_outcome, N_impact - V_study, N_implication - J_important, N_implication - V_understand, N_importance - V_understand, N_incentive - V_create, N_income - N_change, N_income - V_affect, N_income - V_depend, N_income - V_measure, N_increase - N_response, N_increase - V_cause, N_index - V_measure, N_individual - J_different, N_individual - J_particular, N_individual - J_same, N_individual - V_affect, N_individual - V_vary, N_industry - J_such, N_inequality - J_economic, N_influence - J_political, N_information - J_different, N_information - J_important, N_information - J_specific, N_information - J_such, N_insight - J_important, N_institution - J_major, N_institution - J_political, N_institution - J_such, N_institution - V_create, N_instruction - V_follow, N_interaction - V_shape, N_interest - J_particular, N_issue - J_important, N_issue - J_key, N_issue - J_particular, N_issue - J_political, N_issue - J_such, N_issue - V_study, N_item - J_such, N_item - N_response, N_job - V_create, N_kind - J_certain, N_kind - J_different, N_kind - J_particular, N_kind - J_same, N_kind - J_specific, N_kind - J_various, N_knowledge - J_specific, N_language - J_different, N_language - J_particular, N_language - J_specific, N_language - J_such, N_language - V_study, N_language - V_understand, N_law - J_such, N_leader - J_political, N_learning - V_occur, N_level - J_different, N_level - J_economic, N_level - J_individual, N_level - J_same, N_level - N_change, N_level - V_depend, N_level - V_determine, N_level - V_measure, N_level - V_vary, N_life - V_affect, N_life - V_improve, N_life - V_shape, N_location - J_different, N_location - J_particular, N_location - J_specific, N_loss - V_cause, N_market - J_such, N_market - N_outcome, N_market - V_affect,

N_market - V_change, N_market - V_determine, N_material - J_such, N_matter - J_simple, N_meaning - J_different, N_meaning - V_understand, N_measure - J_economic, N_measure - J_such, N_measure - N_outcome, N_measure - V_reflect, N_mechanism - J_different, N_mechanism - J_key, N_mechanism - J_specific, N_member - J_different, N_member - J_individual, N_member - J_same, N_method - J_different, N_method - V_determine, N_method - V_measure, N_method - V_study, N_mind - J_important, N_mind - V_change, N_model - J_different, N_model - J_economic, N_model - J_simple, N_model - J_such, N_model - V_follow, N_model - V_see, N_month - V_follow, N_movement - J_political, N_nation - J_economic, N_nature - V_depend, N_nature - V_reflect, N_nature - V_understand, N_need - J_individual, N_nothing - V_change, N_number - J_different, N_number - J_same, N_number - V_depend, N_number - V_measure, N_opportunity - J_economic, N_opportunity - V_create, N_order - V_determine, N_order - V_follow, N_order - V_understand, N_organization - J_political, N_organization - J_such, N_other - V_see, N_other - V_understand, N_outcome - J_economic, N_outcome - J_such, N_outcome - V_affect, N_outcome - V_improve, N_outcome - V_influence, N_outcome - V_shape, N_page - V_see, N_panel - V_see, N_paper - V_follow, N_parameter - J_key, N_part - J_different, N_part - J_important, N_part - J_major, N_part - J_various, N_participant - J_same, N_participant - V_study, N_party - J_political, N_path - V_follow, N_pattern - J_different, N_pattern - J_same, N_pattern - N_change, N_pattern - V_change, N_pattern - V_create, N_pattern - V_follow, N_pattern - V_reflect, N_pattern - V_see, N_people - J_certain, N_people - J_different, N_people - J_same, N_people - J_such, N_people - V_affect, N_people - V_change, N_people - V_create, N_people - V_see, N_people - V_study, N_people - V_understand, N_percentage - N_change, N_perception - V_affect, N_perception - V_influence, N_perception - V_shape, N_performance - J_individual, N_performance - N_outcome, N_performance - V_affect, N_performance - V_improve, N_performance - V_influence, N_performance - V_measure, N_period - J_different, N_period - J_same, N_period - V_follow, N_period - V_occur, N_person - J_same, N_perspective - V_understand, N_phenomenon - V_occur, N_phenomenon - V_understand, N_place - J_different, N_place - J_particular, N_place - J_such, N_place - N_change, N_player - J_major, N_point - J_different, N_point - J_important, N_point - J_key, N_policy - J_economic, N_policy - J_important, N_policy - J_particular, N_policy - J_such, N_policy - J_various, N_policy - N_change, N_policy - N_response, N_policy - V_affect, N_policy - V_change, N_policy - V_create, N_policy - V_follow, N_policy - V_influence, N_population - J_different, N_population - N_change, N_population - V_occur, N_population - V_study, N_population - V_vary, N_position - J_political, N_power - J_economic, N_power - J_political, N_power - V_create, N_practice - J_such, N_practice - V_follow, N_practice - V_improve, N_practice - V_vary, N_prediction - J_specific, N_preference - J_individual, N_preference - V_reflect, N_pressure - V_create, N_price - J_same, N_price - N_change, N_price - V_affect, N_price - V_cause, N_price - V_change, N_price - V_depend, N_price - V_determine, N_price - V_reflect, N_principle - J_same, N_principle - V_follow, N_principle - V_understand, N_problem - J_economic, N_problem - J_major, N_problem - J_particular, N_problem - J_specific, N_problem - J_such, N_problem - V_cause, N_problem - V_create, N_procedure - J_simple, N_procedure - V_follow, N_process - J_different, N_process - J_political, N_process - J_such, N_process - N_change, N_process - N_outcome, N_process - V_affect, N_process - V_create, N_process - V_follow, N_process - V_influence, N_process - V_occur, N_process - V_reflect, N_process - V_understand, N_product - J_different, N_product - J_particular, N_product - J_same, N_product - J_such, N_product - V_create, N_productivity - V_improve, N_profit - J_economic, N_program - J_such, N_program - V_create, N_progress - J_economic, N_purpose - J_specific, N_quality - V_improve, N_quality - V_measure, N_quality - V_vary, N_question - J_different, N_question - J_important, N_question - J_key, N_question - J_simple, N_question - J_specific, N_question - J_such, N_question - N_response, N_race - J_different, N_range - N_outcome, N_rate - J_different, N_rate - J_same, N_rate - N_change, N_rate - N_response, N_rate - V_affect, N_rate - V_cause, N_rate - V_depend, N_rate - V_determine, N_ratio - V_measure, N_reason - J_different, N_reason - J_important, N_reason - J_major, N_reason - J_same, N_reason - J_simple, N_reason - J_various, N_recession - J_economic, N_reform - J_economic, N_region - J_different, N_region - J_particular, N_region - J_same, N_region - J_specific, N_region - V_vary, N_relationship - V_study, N_relationship - V_understand, N_research - J_important, N_research - J_such, N_researcher - V_measure, N_researcher - V_study, N_researcher - V_understand, N_resource - J_economic, N_resource - J_political, N_resource - J_such, N_response - N_change, N_response - V_change, N_response - V_occur, N_result - J_different, N_result - J_same, N_result - N_change, N_result - V_change, N_review - V_see, N_risk - V_affect, N_role - J_economic, N_role -

J_important, N_role - J_key, N_role - J_major, N_role - J_political, N_role - V_determine, N_role - V_shape, N_role - V_understand, N_rule - J_simple, N_rule - V_change, N_rule - V_follow, N_scale - V_measure, N_science - J_political, N_scientist - J_political, N_scientist - V_study, N_section - V_see, N_selection - V_occur, N_service - J_such, N_set - J_different, N_set - J_particular, N_set - J_same, N_set - J_specific, N_setting - J_different, N_setting - J_particular, N_sex - J_same, N_shape - V_determine, N_shift - V_cause, N_shift - V_occur, N_shift - V_reflect, N_shock - N_response, N_shock - V_follow, N_site - J_important, N_site - J_particular, N_situation - J_certain, N_situation - J_different, N_situation - J_particular, N_situation - J_specific, N_situation - J_such, N_situation - V_create, N_situation - V_depend, N_situation - V_improve, N_situation - V_occur, N_size - J_different, N_size - J_same, N_size - N_change, N_size - V_depend, N_size - V_determine, N_size - V_influence, N_size - V_vary, N_skill - J_specific, N_society - J_different, N_society - J_economic, N_society - J_particular, N_society - J_such, N_society - V_study, N_society - V_vary, N_source - J_different, N_source - J_important, N_source - J_major, N_source - J_various, N_specie - J_different, N_species - J_same, N_stage - J_different, N_stage - J_various, N_state - J_economic, N_state - J_political, N_state - J_such, N_state - V_vary, N_status - J_economic, N_status - V_reflect, N_step - V_follow, N_stimulus - J_different, N_stimulus - J_specific, N_stimulus - J_such, N_stimulus - N_response, N_strategy - J_different, N_strategy - V_follow, N_stress - N_response, N_stress - V_cause, N_structure - J_different, N_structure - J_political, N_structure - J_simple, N_structure - N_change, N_structure - V_affect, N_structure - V_change, N_structure - V_create, N_structure - V_determine, N_structure - V_reflect, N_structure - V_shape, N_student - V_study, N_study - J_such, N_style - J_different, N_subject - V_see, N_success - J_economic, N_success - V_depend, N_supply - N_change, N_supply - V_affect, N_support - J_political, N_survey - N_response, N_survey - V_measure, N_system - J_different, N_system - J_economic, N_system - J_political, N_system - J_such, N_system - V_create, N_task - J_different, N_task - J_particular, N_task - J_simple, N_task - J_specific, N_task - J_such, N_technique - J_various, N_technique - V_measure, N_technology - N_change, N_technology - V_change, N_technology - V_improve, N_term - J_different, N_term - J_economic, N_term - J_key, N_term - J_same, N_term - V_measure, N_term - V_understand, N_test - V_measure, N_theory - J_economic, N_thing - J_certain, N_thing - J_different, N_thing - J_important, N_thing - J_same, N_thing - J_simple, N_thing - J_such, N_thing - V_change, N_time - J_different, N_time - J_particular, N_time - J_same, N_time - J_such, N_time - N_change, N_time - N_response, N_time - V_change, N_time - V_occur, N_time - V_vary, N_tool - J_simple, N_tool - J_such, N_topic - J_important, N_topic - V_study, N_trait - J_certain, N_trait - J_specific, N_trait - J_such, N_transaction - J_economic, N_treatment - J_specific, N_trend - V_affect, N_trend - V_reflect, N_turn - V_affect, N_turn - V_cause, N_turn - V_depend, N_type - J_certain, N_type - J_different, N_type - J_particular, N_type - J_specific, N_type - J_various, N_type - V_depend, N_type - V_vary, N_understanding - V_improve, N_unit - J_political, N_unit - V_measure, N_use - J_such, N_value - J_different, N_value - J_economic, N_value - V_create, N_value - V_depend, N_value - V_determine, N_variable - J_different, N_variable - J_important, N_variable - J_key, N_variable - J_such, N_variable - N_change, N_variable - N_outcome, N_variable - N_response, N_variable - V_affect, N_variable - V_change, N_variable - V_depend, N_variable - V_determine, N_variable - V_influence, N_variable - V_measure, N_variable - V_study, N_variation - J_individual, N_variation - V_reflect, N_variety - J_different, N_version - J_simple, N_view - J_different, N_violence - J_political, N_wage - N_change, N_wage - V_affect, N_way - J_certain, N_way - J_different, N_way - J_important, N_way - J_particular, N_way - J_same, N_way - J_simple, N_way - J_various, N_way - V_affect, N_way - V_change, N_way - V_improve, N_way - V_influence, N_way - V_measure, N_way - V_see, N_way - V_shape, N_way - V_study, N_way - V_understand, N_welfare - V_measure, N_word - J_different, N_word - J_same, N_word - J_such, N_world - J_different, N_world - J_major, N_world - V_change, N_world - V_see, N_world - V_understand, N_year - J_same, V_accord - V_vary, V_achieve - N_outcome, V_adapt - V_change, V_affect - J_economic, V_affect - N_change, V_affect - N_outcome, V_apply - J_same, V_associate - J_different, V_associate - J_particular, V_associate - J_specific, V_associate - N_change, V_associate - N_outcome, V_associate - N_response, V_attempt - V_understand, V_base - J_different, V_become - J_important, V_become - J_major, V_begin - V_understand, V_belong - J_same, V_bring - N_change, V_cause - N_change, V_change - J_economic, V_change - N_response, V_change - V_change, V_come - J_different, V_come - V_understand, V_compare - J_different, V_compare - N_outcome, V_consider - J_important, V_consider - J_simple, V_create - J_economic, V_depend - J_different, V_depend - J_particular,

V_depend - J_specific, V_depend - V_vary, V_desire - N_outcome, V_determine - J_important, V_differ - V_depend, V_discuss - J_various, V_drive - N_change, V_evolve - N_response, V_examine - J_various, V_exist - J_such, V_expect - N_outcome, V_experience - N_change, V_explain - J_certain, V_find - J_different, V_find - J_same, V_find - J_such, V_focus - J_individual, V_focus - J_particular, V_focus - J_specific, V_focus - N_outcome, V_follow - N_change, V_gain - J_political, V_give - J_different, V_give - J_same, V_happen - V_see, V_help - V_create, V_help - V_determine, V_help - V_improve, V_help - V_understand, V_identify - J_key, V_identify - J_specific, V_illustrate - J_key, V_improve - J_economic, V_improve - N_outcome, V_include - J_such, V_include - J_various, V_increase - J_economic, V_influence - J_individual, V_influence - J_political, V_influence - N_outcome, V_involve - J_different, V_involve - N_change, V_keep - J_important, V_keep - J_simple, V_lead - J_different, V_lead - J_economic, V_lead - N_change, V_lead - N_outcome, V_learn - V_study, V_live - J_same, V_look - J_different, V_look - V_see, V_maintain - V_create, V_make - J_different, V_make - J_important, V_make - J_same, V_make - J_such, V_make - N_change, V_measure - N_change, V_meet - J_certain, V_need - V_determine, V_need - V_understand, V_note - J_important, V_occur - J_such, V_occur - N_change, V_occur - N_response, V_operate - J_different, V_play - J_important, V_play - J_key, V_play - J_major, V_predict - N_change, V_predict - N_outcome, V_produce - J_different, V_produce - J_same, V_produce - N_change, V_produce - N_outcome, V_promote - J_economic, V_propose - J_various, V_provide - J_important, V_provide - J_such, V_recognize - J_important, V_reflect - J_different, V_reflect - N_change, V_remain - J_same, V_remember - J_important, V_report - N_outcome, V_represent - J_different, V_respond - N_change, V_result - N_change, V_see - N_change, V_seek - V_understand, V_shape - J_political, V_shape - N_outcome, V_share - J_same, V_show - J_different, V_show - J_same, V_show - N_change, V_show - N_response, V_spend - V_study, V_stay - J_same, V_summarize - J_key, V_take - J_different, V_take - N_change, V_try - V_change, V_try - V_determine, V_try - V_understand, V_understand - J_important, V_use - J_different, V_use - J_same, V_use - J_simple, V_use - J_such, V_use - V_determine, V_use - V_measure, V_use - V_study, V_vary - J_different, V_vary - V_depend, V_want - V_understand
Collocational Dimension 5: Empirical Research Methods (1,069 collocations)
J_absolute - N_term, J_accurate - N_estimate, J_actual - N_difference, J_additional - N_analysis, J_additional - N_evidence, J_additional - N_variable, J_alternative - N_measure, J_alternative - N_method, J_appropriate - N_method, J_available - N_datum, J_available - N_evidence, J_available - N_measure, J_average - N_difference, J_average - N_effect, J_basic - N_term, J_behavioral - N_difference, J_biological - N_difference, J_causal - N_effect, J_causal - N_relationship, J_classic - N_example, J_classic - N_study, J_clear - N_evidence, J_close - N_relationship, J_comparative - N_analysis, J_comparative - N_study, J_complex - N_relationship, J_considerable - N_evidence, J_consistent - N_datum, J_consistent - N_estimate, J_consistent - N_evidence, J_consistent - N_model, J_consistent - N_result, J_constant - N_variable, J_continuous - N_variable, J_cross-sectional - N_datum, J_cultural - N_difference, J_dependent - N_measure, J_dependent - N_variable, J_detailed - N_analysis, J_detailed - N_datum, J_different - J_similar, J_different - N_effect, J_different - N_estimate, J_different - N_method, J_different - N_model, J_different - N_result, J_different - N_term, J_different - N_variable, J_differential - N_effect, J_direct - N_effect, J_direct - N_evidence, J_direct - N_measure, J_due - N_difference, J_due - N_effect, J_dynamic - N_model, J_early - N_evidence, J_early - N_study, J_economic - N_difference, J_economic - N_measure, J_economic - N_model, J_economic - N_term, J_emotional - J_negative, J_emotional - J_positive, J_empirical - N_analysis, J_empirical - N_datum, J_empirical - N_evidence, J_empirical - N_result, J_empirical - N_study, J_empirical - N_test, J_entire - N_sample, J_estimated - N_effect, J_ethnic - N_difference, J_ethnographic - N_datum, J_ethnographic - N_study, J_experimental - N_evidence, J_experimental - N_method, J_experimental - N_study, J_explanatory - N_variable, J_few - N_study, J_first - N_difference, J_first - N_equation, J_first - N_evidence, J_first - N_experiment, J_first - N_study, J_first - N_table, J_first - N_term, J_following - N_equation, J_full - N_sample, J_further - N_analysis, J_further - N_evidence, J_future - N_study, J_general - N_model, J_general - N_term, J_genetic - N_difference, J_genetic - N_evidence, J_genetic - N_study, J_good - N_estimate, J_good - N_evidence, J_good - N_example, J_good - N_measure, J_good - N_model, J_good - N_result, J_great - N_difference, J_great - N_effect, J_historical - N_datum, J_historical - N_evidence, J_human - N_difference, J_human - N_evidence, J_human - N_study, J_important - N_difference, J_important - N_variable, J_independent -

N_effect, J_independent - N_variable, J_indirect - N_effect, J_indirect - N_evidence, J_indirect - N_measure, J_individual - N_difference, J_interesting - N_result, J_key - N_difference, J_key - N_term, J_key - N_variable, J_large - N_difference, J_large - N_effect, J_large - N_sample, J_large - N_study, J_last - N_table, J_least - N_estimate, J_least - N_regression, J_linear - N_model, J_linear - N_regression, J_linear - N_relationship, J_linear - V_estimate, J_little - N_difference, J_little - N_evidence, J_long-term - N_relationship, J_long - N_term, J_main - N_analysis, J_main - N_difference, J_main - N_effect, J_main - N_result, J_many - N_example, J_many - N_study, J_many - N_variable, J_mean - N_difference, J_more - N_datum, J_more - N_study, J_more - N_variable, J_most - N_study, J_multiple - N_analysis, J_multiple - N_method, J_multiple - N_model, J_multiple - N_variable, J_natural - N_experiment, J_negative - J_positive, J_negative - J_significant, J_negative - N_effect, J_negative - N_estimate, J_negative - N_relationship, J_net - N_result, J_new - N_evidence, J_new - N_method, J_new - N_model, J_null - N_result, J_observed - N_datum, J_observed - N_variable, J_online - N_table, J_only - N_difference, J_ordinary - N_regression, J_original - N_study, J_other - J_similar, J_other - N_study, J_other - N_variable, J_overall - J_significant, J_overall - N_effect, J_own - N_term, J_personal - N_relationship, J_physical - N_difference, J_positive - J_negative, J_positive - J_significant, J_positive - N_effect, J_positive - N_relationship, J_possible - N_effect, J_precise - N_estimate, J_present - N_study, J_previous - N_example, J_previous - N_study, J_racial - N_difference, J_random - N_effect, J_random - N_model, J_random - N_sample, J_random - N_variable, J_real - N_difference, J_real - N_term, J_recent - N_datum, J_recent - N_evidence, J_recent - N_example, J_recent - N_study, J_regional - N_difference, J_robust - N_result, J_same - N_participant, J_same - N_result, J_same - N_term, J_scientific - N_evidence, J_scientific - N_method, J_scientific - N_study, J_second - N_experiment, J_second - N_term, J_second - N_variable, J_several - N_method, J_several - N_study, J_sexual - N_difference, J_sexual - N_relationship, J_short - N_term, J_significant - J_negative, J_significant - J_positive, J_significant - N_difference, J_significant - N_effect, J_significant - N_estimate, J_significant - N_relationship, J_significant - N_result, J_significant - N_variable, J_similar - N_effect, J_similar - N_estimate, J_similar - N_experiment, J_similar - N_result, J_similar - N_study, J_simple - N_analysis, J_simple - N_effect, J_simple - N_example, J_simple - N_model, J_single - N_variable, J_small - J_significant, J_small - N_difference, J_small - N_effect, J_small - N_sample, J_small - N_study, J_social - N_relationship, J_standard - N_estimate, J_standard - N_measure, J_standard - N_model, J_statistical - N_analysis, J_statistical - N_datum, J_statistical - N_method, J_statistical - N_test, J_strong - J_negative, J_strong - J_positive, J_strong - N_effect, J_strong - N_evidence, J_strong - N_relationship, J_structural - N_equation, J_substantial - N_difference, J_substantial - N_evidence, J_such - N_datum, J_such - N_difference, J_such - N_effect, J_such - N_evidence, J_such - N_measure, J_such - N_model, J_such - N_study, J_such - N_variable, J_theoretical - N_model, J_third - N_variable, J_total - N_sample, J_traditional - N_method, J_useful - N_analysis, N_adult - N_sample, N_advantage - J_significant, N_age - N_difference, N_age - N_relationship, N_age - N_sample, N_amount - J_significant, N_analysis - N_datum, N_analysis - N_effect, N_analysis - N_method, N_analysis - N_regression, N_analysis - N_result, N_analysis - N_sample, N_analysis - N_variable, N_anthropology - N_study, N_approach - J_similar, N_argument - J_similar, N_aspect - J_positive, N_association - J_negative, N_association - J_positive, N_association - J_significant, N_assumption - N_model, N_background - J_similar, N_basis - N_evidence, N_behavior - N_study, N_bias - J_significant, N_body - N_evidence, N_business - N_model, N_case - N_study, N_cause - N_effect, N_change - J_significant, N_change - N_effect, N_change - N_result, N_change - N_variable, N_characteristic - J_similar, N_child - N_study, N_coefficient - J_negative, N_coefficient - J_positive, N_coefficient - J_significant, N_coefficient - N_equation, N_coefficient - N_estimate, N_coefficient - N_model, N_coefficient - N_regression, N_coefficient - N_table, N_coefficient - N_variable, N_coefficient - V_estimate, N_collection - N_datum, N_column - N_estimate, N_column - N_table, N_comparison - N_analysis, N_conclusion - J_similar, N_conclusion - N_evidence, N_condition - J_similar, N_condition - N_participant, N_consequence - J_negative, N_control - N_participant, N_control - N_variable, N_correlation - J_negative, N_correlation - J_positive, N_correlation - J_significant, N_correlation - N_measure, N_correlation - N_variable, N_cost - V_estimate, N_data - N_analysis, N_datum - N_analysis, N_datum - N_method, N_datum - N_model, N_datum - N_sample, N_datum - N_study, N_datum - N_table, N_datum - V_estimate, N_decline - J_significant, N_design - N_experiment, N_difference - J_significant, N_difference - N_effect, N_difference - N_result, N_difference - N_test, N_display - N_table, N_economy - N_model, N_effect - J_negative, N_effect - J_positive, N_effect - J_significant,

N_effect - J_similar, N_effect - N_analysis, N_effect - N_difference, N_effect - N_estimate, N_effect - N_evidence, N_effect - N_model, N_effect - N_result, N_effect - N_study, N_effect - N_variable, N_effect - V_estimate, N_emotion - J_negative, N_emotion - J_positive, N_equation - N_model, N_equation - N_regression, N_equation - N_term, N_equation - V_estimate, N_equilibrium - N_model, N_error - N_estimate, N_error - N_model, N_error - N_term, N_error - V_estimate, N_estimate - J_negative, N_estimate - J_significant, N_estimate - J_similar, N_estimate - N_effect, N_estimate - N_model, N_estimate - N_regression, N_estimate - N_table, N_event - J_negative, N_evidence - N_effect, N_evidence - N_study, N_example - N_study, N_exchange - N_relationship, N_experience - J_negative, N_experience - J_positive, N_experiment - N_participant, N_experiment - N_result, N_factor - N_analysis, N_family - N_relationship, N_favor - N_evidence, N_feature - J_similar, N_feedback - J_positive, N_feeling - J_negative, N_feeling - J_positive, N_field - N_experiment, N_field - N_study, N_finding - J_positive, N_finding - J_similar, N_finding - N_evidence, N_finding - N_study, N_firm - N_example, N_fit - N_model, N_following - N_example, N_function - N_variable, N_gender - N_difference, N_gender - N_effect, N_group - N_difference, N_group - N_participant, N_group - N_relationship, N_group - N_study, N_growth - N_model, N_heterogeneity - N_effect, N_hypothesis - N_datum, N_hypothesis - N_evidence, N_hypothesis - N_test, N_impact - J_negative, N_impact - J_positive, N_impact - J_significant, N_income - N_difference, N_income - N_effect, N_increase - J_significant, N_indicator - N_variable, N_individual - N_relationship, N_individual - N_sample, N_information - N_datum, N_intelligence - N_test, N_interaction - J_positive, N_interaction - J_significant, N_interaction - N_effect, N_interaction - N_model, N_interaction - N_term, N_interest - N_variable, N_interpretation - N_result, N_intervention - N_effect, N_kind - N_evidence, N_knowledge - N_method, N_laboratory - N_experiment, N_level - J_significant, N_level - N_analysis, N_level - N_variable, N_likelihood - V_estimate, N_line - N_evidence, N_list - N_table, N_magnitude - J_similar, N_magnitude - N_effect, N_magnitude - N_estimate, N_market - N_measure, N_market - N_participant, N_mean - N_sample, N_meaning - N_term, N_measure - N_study, N_measure - N_variable, N_memory - N_test, N_method - N_analysis, N_method - N_datum, N_method - N_study, N_method - V_estimate, N_model - N_datum, N_model - N_effect, N_model - N_equation, N_model - N_estimate, N_model - N_regression, N_model - N_result, N_model - N_table, N_model - N_term, N_model - N_variable, N_model - V_estimate, N_moderator - N_effect, N_moderator - N_variable, N_mood - J_negative, N_nature - N_relationship, N_number - N_study, N_number - N_term, N_number - N_variable, N_observation - N_datum, N_observation - N_method, N_observation - N_participant, N_observation - N_sample, N_outcome - J_negative, N_outcome - J_positive, N_outcome - N_difference, N_outcome - N_effect, N_outcome - N_measure, N_outcome - N_variable, N_panel - N_datum, N_panel - N_table, N_parameter - N_equation, N_parameter - N_estimate, N_parameter - N_model, N_parameter - V_estimate, N_participant - N_experiment, N_participant - N_sample, N_participant - N_study, N_partner - N_relationship, N_pattern - J_similar, N_pattern - N_datum, N_pattern - N_result, N_people - N_example, N_performance - N_effect, N_performance - N_measure, N_performance - N_test, N_period - N_sample, N_piece - N_evidence, N_point - N_estimate, N_policy - N_analysis, N_policy - N_effect, N_population - N_difference, N_population - N_estimate, N_population - N_model, N_population - N_sample, N_power - N_relationship, N_power - N_test, N_practice - N_effect, N_prediction - N_model, N_present - N_table, N_price - N_effect, N_probability - N_model, N_problem - J_similar, N_quality - N_measure, N_quality - N_relationship, N_race - N_term, N_rate - N_difference, N_reaction - J_negative, N_reduction - J_significant, N_regression - N_analysis, N_regression - N_equation, N_regression - N_estimate, N_regression - N_model, N_regression - N_result, N_regression - N_table, N_regression - N_variable, N_regression - V_estimate, N_relationship - J_negative, N_relationship - J_positive, N_relationship - J_significant, N_relationship - N_study, N_relationship - N_variable, N_relationship - V_estimate, N_report - N_table, N_research - N_effect, N_research - N_method, N_research - N_participant, N_research - N_result, N_research - N_study, N_researcher - N_datum, N_researcher - N_experiment, N_researcher - N_method, N_researcher - N_study, N_response - J_positive, N_response - N_variable, N_result - J_significant, N_result - J_similar, N_result - N_analysis, N_result - N_difference, N_result - N_effect, N_result - N_experiment, N_result - N_model, N_result - N_regression, N_result - N_study, N_result - N_table, N_risk - J_significant, N_risk - N_measure, N_role - J_significant, N_sample - N_analysis, N_sample - N_datum, N_sample - N_participant, N_sample - N_study, N_sample - V_estimate, N_score - N_test, N_search - N_term, N_sensitivity - N_analysis, N_series - N_analysis, N_series - N_experiment, N_set - N_datum, N_set - N_variable, N_sex -

N_difference, N_show - N_table, N_side - N_equation, N_sign - J_negative, N_sign - J_positive, N_significance - N_test, N_similarity - N_difference, N_situation - J_similar, N_size - N_difference, N_size - N_effect, N_size - N_estimate, N_size - N_relationship, N_size - N_sample, N_size - N_study, N_source - N_datum, N_specification - N_table, N_specification - V_estimate, N_square - N_regression, N_state - N_variable, N_statistic - N_table, N_statistic - N_test, N_status - N_difference, N_status - N_relationship, N_stereotype - J_negative, N_strength - N_relationship, N_structure - N_term, N_student - N_sample, N_student - N_study, N_student - N_test, N_study - J_similar, N_study - N_datum, N_study - N_effect, N_study - N_evidence, N_study - N_example, N_study - N_method, N_study - N_participant, N_study - N_relationship, N_study - N_result, N_study - N_sample, N_subject - N_experiment, N_summary - N_table, N_support - N_evidence, N_survey - N_datum, N_survey - N_result, N_survey - N_sample, N_table - N_estimate, N_table - N_result, N_task - N_participant, N_tax - N_effect, N_term - N_equation, N_test - N_difference, N_testing - N_effect, N_theory - N_evidence, N_theory - N_method, N_time - N_analysis, N_time - N_datum, N_time - N_effect, N_time - N_model, N_time - N_variable, N_treatment - N_effect, N_trial - N_participant, N_type - N_datum, N_unit - N_analysis, N_use - N_datum, N_use - N_measure, N_use - N_method, N_use - N_term, N_value - J_negative, N_value - J_positive, N_value - N_difference, N_value - N_variable, N_value - V_estimate, N_variable - J_significant, N_variable - N_analysis, N_variable - N_effect, N_variable - N_measure, N_variable - N_model, N_variable - N_regression, N_variable - N_relationship, N_variance - N_analysis, N_variation - N_sample, N_variety - N_method, N_version - N_model, N_wage - N_difference, N_wage - N_effect, N_way - J_similar, N_woman - N_study, N_year - N_datum, V_accord - N_model, V_account - N_difference, V_accumulate - N_evidence, V_add - N_model, V_add - N_variable, V_affect - N_variable, V_allow - N_method, V_allow - N_model, V_analyze - N_datum, V_apply - J_similar, V_apply - N_analysis, V_apply - N_datum, V_apply - N_method, V_apply - N_model, V_apply - N_term, V_ask - N_experiment, V_ask - N_participant, V_assess - N_measure, V_assess - N_method, V_assess - N_relationship, V_assess - N_study, V_assign - N_participant, V_associate - J_negative, V_associate - J_positive, V_assume - N_model, V_attribute - N_difference, V_base - N_analysis, V_base - N_datum, V_base - N_difference, V_base - N_estimate, V_base - N_evidence, V_base - N_measure, V_base - N_method, V_base - N_model, V_base - N_result, V_base - N_sample, V_base - N_study, V_base - N_test, V_begin - N_analysis, V_begin - N_study, V_bias - N_estimate, V_build - N_model, V_build - N_relationship, V_calculate - N_estimate, V_call - N_model, V_capture - N_measure, V_cause - N_difference, V_change - N_result, V_change - N_variable, V_collect - N_datum, V_collect - N_sample, V_come - N_datum, V_come - N_evidence, V_come - N_study, V_compare - N_effect, V_compare - N_estimate, V_compare - N_model, V_compare - N_result, V_compare - N_study, V_complete - N_participant, V_compute - N_measure, V_conduct - N_analysis, V_conduct - N_experiment, V_conduct - N_study, V_conduct - N_test, V_confirm - N_evidence, V_confirm - N_result, V_consider - N_effect, V_consider - N_example, V_consider - N_model, V_consider - N_study, V_construct - N_datum, V_construct - N_measure, V_contain - N_datum, V_contain - N_model, V_contain - N_table, V_control - N_experiment, V_control - N_variable, V_correlate - N_measure, V_correlate - N_variable, V_define - N_term, V_define - N_variable, V_demonstrate - N_effect, V_demonstrate - N_evidence, V_demonstrate - N_experiment, V_demonstrate - N_study, V_denote - N_variable, V_depend - N_effect, V_depend - N_variable, V_derive - N_equation, V_describe - N_datum, V_describe - N_equation, V_describe - N_experiment, V_describe - N_method, V_describe - N_model, V_describe - N_relationship, V_describe - N_study, V_describe - N_term, V_design - N_experiment, V_detect - N_effect, V_determine - N_method, V_determine - N_variable, V_develop - N_measure, V_develop - N_method, V_develop - N_model, V_develop - N_relationship, V_differ - N_term, V_discuss - N_method, V_discuss - N_model, V_display - N_table, V_draw - N_sample, V_drive - N_result, V_emerge - N_evidence, V_employ - N_measure, V_employ - N_method, V_engage - N_participant, V_establish - N_relationship, V_estimate - N_datum, V_estimate - N_effect, V_estimate - N_equation, V_estimate - N_method, V_estimate - N_model, V_estimate - N_regression, V_estimate - N_relationship, V_estimate - N_sample, V_examine - N_analysis, V_examine - N_datum, V_examine - N_effect, V_examine - N_relationship, V_examine - N_study, V_exclude - N_analysis, V_exist - J_similar, V_exist - N_difference, V_exist - N_evidence, V_exist - N_relationship, V_expect - J_positive, V_expect - N_effect, V_experience - J_negative, V_experience - J_significant, V_experience - N_participant, V_explain - N_difference, V_explain - N_effect, V_explain - N_model, V_explain - N_result, V_explain - N_term, V_explore - N_relationship, V_express - N_term, V_extend - N_analysis, V_face - J_similar, V_find - J_significant,

V_find - J_similar, V_find - N_difference, V_find - N_effect, V_find - N_evidence, V_find - N_example, V_find - N_relationship, V_find - N_result, V_find - N_study, V_fit - N_datum, V_fit - N_model, V_fix - N_effect, V_fix - N_model, V_focus - N_analysis, V_focus - N_study, V_follow - N_model, V_form - N_relationship, V_gather - N_datum, V_generate - J_negative, V_generate - J_positive, V_generate - N_datum, V_generate - N_model, V_give - N_datum, V_give - N_equation, V_give - N_estimate, V_give - N_example, V_give - N_participant, V_give - N_sample, V_give - N_table, V_identify - N_method, V_identify - N_study, V_illustrate - N_example, V_illustrate - N_method, V_illustrate - N_relationship, V_illustrate - N_table, V_imply - N_model, V_imply - N_term, V_include - N_analysis, V_include - N_datum, V_include - N_effect, V_include - N_example, V_include - N_measure, V_include - N_model, V_include - N_regression, V_include - N_sample, V_include - N_study, V_include - N_table, V_include - N_term, V_include - N_variable, V_increase - N_effect, V_indicate - J_negative, V_indicate - J_positive, V_indicate - J_significant, V_indicate - N_analysis, V_indicate - N_datum, V_indicate - N_estimate, V_indicate - N_evidence, V_indicate - N_participant, V_indicate - N_result, V_indicate - N_study, V_indicate - N_table, V_indicate - N_variable, V_influence - N_variable, V_interpret - N_evidence, V_interpret - N_result, V_introduce - N_model, V_investigate - N_relationship, V_investigate - N_study, V_involve - N_method, V_involve - N_participant, V_involve - N_study, V_lead - J_negative, V_lead - J_positive, V_lead - N_difference, V_learn - N_participant, V_limit - N_sample, V_list - N_table, V_maintain - N_relationship, V_make - N_difference, V_match - N_datum, V_match - N_sample, V_mean - N_difference, V_mean - N_term, V_measure - N_effect, V_measure - N_method, V_measure - N_term, V_measure - N_test, V_measure - N_variable, V_miss - N_datum, V_moderate - N_effect, V_moderate - N_relationship, V_observe - J_significant, V_observe - N_datum, V_observe - N_difference, V_observe - N_effect, V_observe - N_participant, V_obtain - J_similar, V_obtain - N_datum, V_obtain - N_estimate, V_obtain - N_measure, V_obtain - N_result, V_obtain - N_sample, V_occur - J_similar, V_occur - N_effect, V_pass - N_test, V_perform - N_analysis, V_perform - N_experiment, V_perform - N_participant, V_perform - N_test, V_play - J_significant, V_predict - N_difference, V_predict - N_model, V_predict - N_regression, V_present - N_analysis, V_present - N_datum, V_present - N_estimate, V_present - N_evidence, V_present - N_experiment, V_present - N_model, V_present - N_participant, V_present - N_regression, V_present - N_result, V_present - N_table, V_produce - J_similar, V_produce - N_effect, V_produce - N_estimate, V_produce - N_result, V_propose - N_method, V_propose - N_model, V_provide - N_analysis, V_provide - N_datum, V_provide - N_estimate, V_provide - N_evidence, V_provide - N_example, V_provide - N_experiment, V_provide - N_measure, V_provide - N_model, V_provide - N_result, V_provide - N_study, V_provide - N_table, V_provide - N_test, V_publish - N_study, V_receive - N_participant, V_reduce - J_negative, V_refer - N_term, V_reflect - N_difference, V_reflect - N_measure, V_regard - N_evidence, V_relate - N_difference, V_relate - N_variable, V_rely - N_method, V_remain - J_significant, V_report - J_significant, V_report - J_similar, V_report - N_datum, V_report - N_effect, V_report - N_estimate, V_report - N_participant, V_report - N_result, V_report - N_study, V_report - N_table, V_represent - N_variable, V_require - N_analysis, V_require - N_participant, V_respond - N_participant, V_restrict - N_sample, V_reveal - J_significant, V_reveal - N_analysis, V_reveal - N_difference, V_reveal - N_study, V_reveal - N_test, V_review - N_evidence, V_run - N_analysis, V_run - N_regression, V_see - N_effect, V_see - N_equation, V_see - N_example, V_see - N_model, V_see - N_table, V_select - N_sample, V_set - N_datum, V_share - J_similar, V_show - J_positive, V_show - J_significant, V_show - J_similar, V_show - N_analysis, V_show - N_datum, V_show - N_difference, V_show - N_effect, V_show - N_equation, V_show - N_evidence, V_show - N_example, V_show - N_experiment, V_show - N_model, V_show - N_participant, V_show - N_relationship, V_show - N_result, V_show - N_study, V_show - N_table, V_solve - N_equation, V_specify - N_model, V_state - N_term, V_study - N_effect, V_study - N_method, V_study - N_participant, V_study - N_relationship, V_study - N_variable, V_suggest - N_analysis, V_suggest - N_datum, V_suggest - N_difference, V_suggest - N_effect, V_suggest - N_estimate, V_suggest - N_evidence, V_suggest - N_model, V_suggest - N_result, V_suggest - N_study, V_summarize - N_result, V_summarize - N_table, V_support - N_datum, V_support - N_evidence, V_support - N_result, V_support - N_study, V_suppose - N_example, V_take - N_example, V_take - N_measure, V_take - N_sample, V_take - N_test, V_tell - N_participant, V_tend - J_similar, V_test - N_analysis, V_test - N_datum, V_test - N_effect, V_test - N_experiment, V_test - N_model, V_test - N_study, V_think - N_term, V_understand - N_difference, V_understand - N_relationship, V_understand - N_term, V_use - J_similar, V_use - N_analysis, V_use - N_datum, V_use - N_effect, V_use - N_equation, V_use -

N_estimate, V_use - N_example, V_use - N_experiment, V_use - N_measure, V_use - N_method, V_use - N_model, V_use - N_regression, V_use - N_result, V_use - N_sample, V_use - N_study, V_use - N_term, V_use - N_test, V_use - N_variable, V_use - V_estimate, V_yield - J_significant, V_yield - J_similar, V_yield - N_result
Collocational Dimension 6: Local and Global Society (1,109 collocations)
J_active - V_become, J_african - J_american, J_african - N_country, J_american - J_many, J_american - N_city, J_american - N_community, J_american - N_country, J_american - N_family, J_american - N_group, J_american - N_population, J_american - N_society, J_apparent - V_become, J_asian - J_american, J_asian - N_country, J_asian - N_population, J_available - V_become, J_aware - V_become, J_big - J_small, J_big - N_city, J_black - N_community, J_central - N_city, J_certain - N_group, J_clear - V_become, J_close - N_family, J_common - V_become, J_complex - J_large, J_complex - N_organization, J_complex - N_society, J_complex - V_become, J_contemporary - N_society, J_corporate - N_business, J_cultural - J_many, J_cultural - J_other, J_cultural - N_group, J_cultural - N_institution, J_cultural - N_society, J_current - N_state, J_dependent - V_become, J_different - J_many, J_different - J_other, J_different - N_business, J_different - N_country, J_different - N_group, J_different - N_member, J_different - N_part, J_different - N_population, J_different - N_society, J_different - N_world, J_difficult - V_become, J_distinct - N_group, J_diverse - N_group, J_dominant - N_group, J_dominant - V_become, J_due - N_part, J_early - N_part, J_economic - J_many, J_economic - J_other, J_economic - N_country, J_economic - N_society, J_economic - N_state, J_economic - N_world, J_educational - N_institution, J_effective - N_organization, J_emotional - N_state, J_entire - N_population, J_equal - J_other, J_essential - N_part, J_ethnic - N_group, J_european - J_american, J_european - N_country, J_european - N_population, J_expensive - V_become, J_experimental - N_group, J_familiar - V_become, J_federal - N_state, J_few - N_country, J_financial - N_institution, J_first - N_part, J_first - V_become, J_foreign - N_country, J_formal - N_institution, J_formal - N_organization, J_general - N_population, J_global - J_local, J_global - N_economy, J_good - N_business, J_high - N_country, J_high - N_population, J_high - N_state, J_human - J_many, J_human - J_most, J_human - N_animal, J_human - N_body, J_human - N_group, J_human - N_part, J_human - N_population, J_human - N_society, J_human - V_include, J_important - J_many, J_important - N_part, J_important - V_become, J_indian - J_american, J_indigenous - N_population, J_individual - N_group, J_individual - N_member, J_industrial - N_society, J_interested - V_become, J_internal - N_organization, J_international - N_country, J_international - N_organization, J_japanese - J_american, J_key - N_part, J_large - J_small, J_large - N_animal, J_large - N_body, J_large - N_business, J_large - N_city, J_large - N_community, J_large - N_country, J_large - N_economy, J_large - N_family, J_large - N_group, J_large - N_part, J_large - N_population, J_large - N_society, J_large - N_world, J_large - V_become, J_large - V_grow, J_large - V_include, J_large - V_live, J_least - N_part, J_likely - V_become, J_local - N_business, J_local - N_community, J_local - N_economy, J_local - N_group, J_local - N_institution, J_local - N_organization, J_local - N_population, J_local - N_state, J_long - V_live, J_low - N_country, J_major - N_city, J_major - N_institution, J_major - N_part, J_major - N_world, J_major - V_become, J_many - J_american, J_many - J_other, J_many - J_small, J_many - N_animal, J_many - N_business, J_many - N_city, J_many - N_community, J_many - N_country, J_many - N_family, J_many - N_group, J_many - N_part, J_many - N_society, J_many - N_state, J_many - N_world, J_many - V_include, J_many - V_live, J_mental - N_state, J_modern - N_society, J_modern - N_state, J_modern - N_world, J_more - J_large, J_more - J_many, J_more - J_other, J_more - N_country, J_more - N_group, J_more - V_become, J_most - J_other, J_most - N_country, J_most - N_part, J_most - N_society, J_most - N_state, J_most - N_world, J_most - V_live, J_much - N_world, J_national - J_local, J_national - N_organization, J_native - J_american, J_native - N_population, J_natural - N_world, J_new - J_many, J_new - N_business, J_new - N_member, J_new - N_organization, J_new - V_include, J_nuclear - N_family, J_old - V_grow, J_other - J_many, J_other - J_most, J_other - N_animal, J_other - N_business, J_other - N_country, J_other - N_family, J_other - N_group, J_other - N_institution, J_other - N_member, J_other - N_organization, J_other - N_part, J_other - N_population, J_other - N_society, J_other - N_state, J_other - N_world, J_other - V_include, J_outside - N_world, J_overall - N_population, J_own - N_country, J_own - N_family, J_own - N_group, J_own - N_member, J_own - N_society, J_particular - N_country, J_particular - N_group, J_particular - N_society, J_political - N_business, J_political - N_country, J_political - N_economy, J_political - N_group, J_political - N_institution, J_political - N_organization, J_political -

N_society, J_political - N_state, J_poor - N_community, J_poor - N_country, J_poor - N_family, J_poor - N_world, J_popular - V_become, J_possible - J_many, J_professional - N_organization, J_public - N_institution, J_racial - N_group, J_real - N_world, J_regional - J_local, J_religious - N_group, J_rich - N_country, J_rich - V_become, J_rural - N_community, J_rural - V_live, J_same - J_many, J_same - J_other, J_same - N_country, J_same - N_group, J_same - N_member, J_same - V_live, J_scientific - N_community, J_second - N_part, J_several - J_other, J_several - N_country, J_several - N_group, J_several - V_include, J_significant - J_small, J_similar - J_other, J_small - J_large, J_small - J_many, J_small - N_animal, J_small - N_body, J_small - N_business, J_small - N_community, J_small - N_economy, J_small - N_family, J_small - N_group, J_small - N_population, J_small - V_become, J_small - V_include, J_small - V_live, J_social - J_large, J_social - J_many, J_social - J_other, J_social - N_family, J_social - N_group, J_social - N_institution, J_social - N_organization, J_social - N_society, J_social - N_world, J_social - V_include, J_steady - N_state, J_strong - N_state, J_such - J_large, J_such - J_many, J_such - J_other, J_such - N_animal, J_such - N_country, J_such - N_group, J_such - N_institution, J_such - N_organization, J_such - N_society, J_such - N_state, J_such - V_include, J_total - N_population, J_traditional - N_society, J_urban - N_community, J_urban - N_population, J_urban - V_live, J_various - N_country, J_various - N_group, J_various - N_part, J_various - V_include, J_wealthy - N_country, J_western - N_country, J_western - N_society, J_whole - N_society, J_widespread - V_become, J_wild - N_animal, N_activity - J_other, N_activity - N_business, N_age - N_group, N_age - N_population, N_agency - J_local, N_amount - J_large, N_amount - J_small, N_analysis - V_include, N_ancestor - V_live, N_animal - J_large, N_animal - J_many, N_animal - J_other, N_animal - J_small, N_animal - V_include, N_anthropologist - J_many, N_anthropologist - J_most, N_area - J_large, N_area - J_local, N_area - J_many, N_area - J_other, N_area - J_small, N_area - N_city, N_area - N_world, N_area - V_live, N_aspect - J_many, N_aspect - J_other, N_asset - J_other, N_authority - J_local, N_authority - N_state, N_awareness - V_grow, N_band - J_small, N_bank - N_institution, N_behavior - J_other, N_behavior - V_include, N_board - N_member, N_body - J_large, N_body - J_small, N_body - N_part, N_body - V_grow, N_brain - J_large, N_brain - N_body, N_brain - N_part, N_business - J_local, N_business - J_many, N_business - J_other, N_business - J_small, N_business - N_group, N_business - N_organization, N_capital - N_city, N_capital - N_economy, N_case - J_many, N_case - J_most, N_case - J_other, N_category - N_member, N_category - V_include, N_center - N_community, N_change - J_large, N_change - J_small, N_change - N_population, N_characteristic - J_other, N_child - J_many, N_child - J_other, N_child - N_family, N_child - V_grow, N_child - V_include, N_child - V_live, N_city - J_american, N_city - J_large, N_city - J_many, N_city - N_population, N_city - N_state, N_city - V_live, N_class - N_member, N_community - J_american, N_community - J_large, N_community - J_local, N_community - J_many, N_community - J_small, N_community - N_family, N_community - N_member, N_community - N_organization, N_community - V_live, N_company - J_large, N_company - J_many, N_company - J_most, N_company - J_other, N_company - J_small, N_company - V_grow, N_comparison - N_group, N_concern - V_grow, N_condition - J_local, N_condition - J_other, N_condition - V_include, N_condition - V_live, N_control - N_group, N_control - N_state, N_control - V_include, N_corporation - J_large, N_corporation - N_business, N_cost - J_large, N_cost - N_business, N_cost - V_include, N_country - J_american, N_country - J_large, N_country - J_many, N_country - J_most, N_country - J_other, N_country - N_city, N_country - N_economy, N_country - N_part, N_country - N_population, N_country - N_world, N_country - V_include, N_country - V_live, N_crop - V_grow, N_culture - J_american, N_culture - J_local, N_culture - J_many, N_culture - J_other, N_culture - N_group, N_culture - N_member, N_culture - N_part, N_culture - N_society, N_culture - N_world, N_culture - V_include, N_cycle - N_business, N_datum - V_include, N_decision - N_business, N_demand - V_grow, N_density - N_population, N_development - N_organization, N_difference - J_large, N_difference - J_small, N_difference - N_group, N_difference - N_population, N_distribution - N_population, N_economist - J_many, N_economy - J_large, N_economy - J_local, N_economy - J_small, N_economy - N_country, N_economy - N_part, N_economy - N_society, N_economy - N_state, N_economy - N_world, N_economy - V_grow, N_education - N_family, N_effect - J_large, N_effect - J_small, N_effect - V_include, N_environment - J_local, N_environment - V_live, N_error - J_large, N_estimate - N_population, N_evidence - N_body, N_example - J_many, N_example - V_include, N_executive - N_business, N_face - N_body, N_fact - J_many, N_factor - J_many, N_factor - J_other, N_factor - V_include, N_family - J_american, N_family - J_large, N_family - J_many, N_family - J_other, N_family - J_small, N_family - N_community,

N_family - N_group, N_family - N_member, N_family - N_society, N_family - V_include, N_family - V_live, N_farmer - J_local, N_feature - J_many, N_feature - J_other, N_feature - V_include, N_female - N_group, N_firm - J_large, N_firm - J_many, N_firm - J_most, N_firm - J_other, N_firm - J_small, N_firm - N_business, N_following - V_include, N_food - J_other, N_food - N_animal, N_food - V_become, N_food - V_grow, N_food - V_include, N_form - J_many, N_form - J_other, N_form - N_family, N_form - N_organization, N_form - V_include, N_fraction - J_large, N_fraction - J_small, N_fraction - N_population, N_friend - N_family, N_friend - N_member, N_game - J_large, N_generation - J_many, N_good - J_other, N_government - J_local, N_government - J_many, N_government - N_business, N_government - N_economy, N_government - N_institution, N_government - N_state, N_group - J_american, N_group - J_large, N_group - J_local, N_group - J_many, N_group - J_other, N_group - J_small, N_group - N_family, N_group - N_member, N_group - N_organization, N_group - N_population, N_group - N_society, N_group - V_include, N_group - V_live, N_growth - N_economy, N_growth - N_population, N_hand - J_other, N_health - N_community, N_health - N_family, N_health - N_population, N_health - V_include, N_history - J_american, N_history - J_most, N_history - N_family, N_home - N_community, N_home - N_country, N_home - N_family, N_home - V_live, N_host - N_country, N_house - V_live, N_household - N_family, N_household - N_member, N_household - V_live, N_human - J_other, N_human - N_animal, N_human - V_include, N_human - V_live, N_identity - N_group, N_immigrant - N_country, N_impact - J_large, N_income - J_large, N_income - N_family, N_income - V_include, N_increase - J_large, N_increase - J_small, N_increase - N_population, N_individual - J_many, N_individual - J_other, N_individual - J_small, N_individual - N_community, N_individual - N_family, N_individual - N_group, N_individual - N_organization, N_individual - N_population, N_individual - N_society, N_individual - V_become, N_individual - V_live, N_industry - N_business, N_inequality - N_society, N_inequality - V_grow, N_information - J_other, N_information - V_include, N_institution - J_local, N_institution - J_other, N_institution - N_society, N_institution - N_state, N_interest - N_business, N_interest - N_group, N_investment - J_large, N_island - J_small, N_item - V_include, N_kind - J_many, N_kind - J_other, N_knowledge - J_local, N_knowledge - N_body, N_knowledge - N_world, N_language - J_local, N_language - J_other, N_language - N_world, N_language - V_include, N_law - N_state, N_leader - N_business, N_leader - V_become, N_level - J_local, N_level - N_country, N_level - N_group, N_level - N_state, N_life - J_many, N_life - J_other, N_life - N_animal, N_life - N_community, N_life - N_family, N_life - N_part, N_life - V_live, N_literature - J_large, N_literature - N_body, N_literature - V_grow, N_loss - J_large, N_majority - N_population, N_male - J_other, N_male - N_group, N_man - J_american, N_man - J_many, N_man - J_most, N_man - N_society, N_man - V_live, N_manager - N_business, N_market - J_large, N_market - J_local, N_market - J_many, N_market - N_country, N_market - N_economy, N_market - N_world, N_marriage - N_family, N_material - V_include, N_mean - N_population, N_measure - V_include, N_member - J_other, N_member - N_community, N_member - N_family, N_member - N_group, N_member - N_organization, N_member - N_population, N_member - N_society, N_member - V_become, N_membership - N_group, N_mind - N_body, N_minority - N_community, N_minority - N_group, N_minority - N_member, N_model - N_business, N_model - N_economy, N_model - N_population, N_model - V_include, N_money - N_business, N_mother - N_family, N_nation - J_large, N_nation - J_many, N_nation - J_most, N_nation - J_other, N_nation - N_group, N_nation - N_world, N_neighborhood - V_live, N_network - J_local, N_number - J_large, N_number - J_other, N_number - J_small, N_number - N_country, N_number - N_group, N_number - V_grow, N_number - V_include, N_object - J_other, N_official - J_local, N_one - J_small, N_organism - V_live, N_organization - J_local, N_organization - J_other, N_organization - N_business, N_organization - N_community, N_organization - N_group, N_organization - N_member, N_organization - N_society, N_organization - N_state, N_other - J_many, N_owner - N_business, N_pair - N_member, N_parent - N_family, N_parent - V_live, N_part - J_large, N_part - J_many, N_part - J_most, N_part - J_other, N_part - N_body, N_part - N_country, N_part - N_economy, N_part - N_world, N_part - V_become, N_participant - N_group, N_people - J_large, N_people - J_local, N_people - J_many, N_people - J_most, N_people - J_other, N_people - N_community, N_people - N_country, N_people - N_group, N_people - N_society, N_people - N_world, N_people - V_become, N_people - V_include, N_people - V_live, N_percent - N_population, N_percent - N_world, N_percent - V_grow, N_percentage - J_large, N_percentage - J_small, N_percentage - N_population, N_person - J_other, N_place - J_many, N_place - J_other, N_place - N_world, N_place - V_live, N_plant - N_animal, N_plant - V_grow, N_policy - N_economy, N_policy - N_state, N_population - J_american, N_population - J_large, N_population -

J_local, N_population - J_other, N_population - J_small, N_population - N_city, N_population - N_country, N_population - N_group, N_population - N_member, N_population - N_world, N_population - V_become, N_population - V_grow, N_population - V_live, N_portion - J_large, N_poverty - V_live, N_power - N_state, N_price - J_large, N_primate - J_other, N_primate - V_include, N_problem - J_many, N_problem - J_other, N_problem - V_become, N_process - N_part, N_product - J_many, N_product - J_other, N_product - V_include, N_production - N_economy, N_proportion - J_large, N_proportion - J_small, N_proportion - N_population, N_quantity - J_large, N_quantity - J_small, N_question - J_many, N_range - V_include, N_rate - N_country, N_rate - N_economy, N_rate - N_population, N_rate - N_world, N_rate - V_grow, N_reason - J_many, N_reason - J_other, N_region - J_other, N_region - N_country, N_region - N_world, N_region - V_live, N_regression - V_include, N_relationship - N_family, N_relationship - N_group, N_religion - N_world, N_research - J_most, N_research - J_other, N_research - N_animal, N_research - N_body, N_research - V_include, N_researcher - J_many, N_researcher - J_most, N_researcher - J_other, N_resident - J_local, N_resident - N_city, N_resident - N_community, N_resource - J_other, N_resource - N_business, N_resource - N_society, N_resource - V_include, N_response - J_local, N_responsibility - N_family, N_rest - N_world, N_risk - N_business, N_role - N_society, N_room - V_live, N_sample - J_large, N_sample - J_small, N_sample - N_population, N_sample - V_include, N_scale - J_large, N_scale - J_small, N_scale - N_economy, N_scholar - J_many, N_school - J_local, N_scientist - J_many, N_scientist - J_most, N_sector - N_economy, N_segment - N_population, N_sense - N_community, N_set - J_small, N_set - V_include, N_sex - N_member, N_shape - N_body, N_share - J_large, N_share - J_small, N_share - N_population, N_side - J_other, N_site - J_many, N_site - J_other, N_size - J_large, N_size - J_small, N_size - N_body, N_size - N_family, N_size - N_group, N_size - N_population, N_size - V_grow, N_society - J_american, N_society - J_large, N_society - J_many, N_society - J_most, N_society - J_other, N_society - N_economy, N_society - N_family, N_society - N_group, N_society - N_institution, N_society - N_member, N_society - N_organization, N_society - N_state, N_society - N_world, N_society - V_become, N_society - V_live, N_source - J_other, N_specie - J_many, N_specie - J_other, N_specie - N_animal, N_specie - N_member, N_specie - V_include, N_specification - V_include, N_standard - V_live, N_state - J_local, N_state - J_many, N_state - J_most, N_state - J_other, N_state - N_city, N_state - N_economy, N_state - N_institution, N_state - N_organization, N_state - N_society, N_state - N_world, N_state - V_include, N_stock - J_large, N_story - N_part, N_street - N_city, N_structure - N_family, N_structure - N_institution, N_structure - N_organization, N_student - J_american, N_student - J_many, N_student - J_most, N_student - N_group, N_study - J_large, N_study - J_many, N_study - J_most, N_study - J_other, N_study - J_small, N_study - N_group, N_study - V_include, N_subject - N_group, N_support - N_family, N_system - J_other, N_system - N_part, N_system - N_society, N_system - N_world, N_system - V_include, N_target - N_group, N_tax - N_state, N_team - N_member, N_temperature - N_body, N_term - V_include, N_thing - J_many, N_thing - J_other, N_thing - V_live, N_thousand - J_many, N_time - J_large, N_time - J_many, N_time - J_most, N_time - J_other, N_time - V_become, N_time - V_include, N_today - J_many, N_today - J_most, N_today - N_world, N_today - V_live, N_tool - V_include, N_town - J_small, N_town - N_city, N_trade - N_country, N_treatment - N_group, N_treatment - V_include, N_tree - N_family, N_tree - V_grow, N_type - J_many, N_type - J_other, N_type - N_organization, N_type - V_include, N_union - N_member, N_unit - J_large, N_unit - J_small, N_unit - N_family, N_value - J_large, N_value - J_small, N_variable - J_many, N_variable - J_other, N_variable - N_state, N_variable - V_include, N_variation - N_group, N_variety - N_animal, N_variety - V_include, N_village - J_small, N_water - N_body, N_way - J_many, N_way - J_other, N_way - N_society, N_weight - N_body, N_welfare - N_state, N_whole - N_economy, N_whole - N_society, N_woman - J_american, N_woman - J_many, N_woman - J_most, N_woman - N_family, N_woman - N_society, N_woman - V_become, N_woman - V_live, N_word - J_other, N_work - J_other, N_work - N_family, N_work - N_group, N_work - N_organization, N_work - V_include, N_worker - J_many, N_worker - N_country, N_world - J_large, N_world - J_many, N_world - J_most, N_world - J_other, N_world - N_country, N_world - N_economy, N_world - N_part, N_world - N_population, N_world - V_become, N_world - V_live, N_year - J_many, N_year - V_grow, N_year - V_live, V_affect - J_other, V_affect - N_economy, V_agree - J_most, V_base - N_economy, V_base - N_group, V_bear - N_family, V_become - J_large, V_become - J_small, V_become - N_member, V_become - N_part, V_become - N_population, V_become - N_society, V_become - N_world, V_believe - J_many, V_believe - J_most, V_belong - N_group, V_call - N_group, V_change - N_world, V_compare - J_other, V_compare - N_group, V_continue -

V_grow, V_control - N_state, V_cover - N_body, V_create - N_institution, V_define - N_group, V_develop - N_country, V_develop - N_economy, V_develop - N_state, V_develop - N_world, V_differ - N_country, V_differ - N_group, V_eat - N_animal, V_establish - N_institution, V_exist - J_many, V_exist - N_society, V_expect - V_grow, V_experience - N_country, V_find - J_large, V_find - J_many, V_find - J_most, V_find - J_other, V_find - J_small, V_find - N_society, V_fix - V_include, V_form - N_group, V_form - N_part, V_get - J_large, V_go - N_business, V_grow - J_large, V_grow - N_body, V_grow - N_economy, V_grow - N_population, V_help - N_organization, V_hold - J_other, V_hunt - N_animal, V_include - J_large, V_include - J_many, V_include - J_other, V_include - J_small, V_include - N_animal, V_include - N_country, V_include - N_family, V_include - N_group, V_include - N_society, V_include - N_state, V_include - N_world, V_increase - N_population, V_involve - J_many, V_involve - V_become, V_join - N_group, V_kill - N_animal, V_know - J_other, V_know - V_become, V_lead - J_large, V_lead - J_many, V_lead - N_world, V_live - J_large, V_live - J_many, V_live - J_most, V_live - J_small, V_live - N_city, V_live - N_community, V_live - N_country, V_live - N_family, V_live - N_group, V_live - N_population, V_live - N_society, V_live - N_world, V_locate - N_city, V_make - J_large, V_make - J_many, V_make - J_other, V_make - N_business, V_make - N_group, V_make - N_world, V_move - N_city, V_occur - N_population, V_operate - N_business, V_operate - N_country, V_organize - N_group, V_organize - N_society, V_own - N_business, V_participate - N_community, V_play - N_part, V_produce - J_large, V_produce - J_many, V_produce - N_country, V_produce - N_economy, V_relate - J_other, V_remain - J_most, V_represent - N_population, V_require - J_large, V_see - J_many, V_see - J_other, V_see - N_group, V_see - N_world, V_sell - N_business, V_share - J_many, V_share - N_member, V_show - J_other, V_show - N_group, V_spend - J_most, V_start - N_business, V_study - N_community, V_study - N_population, V_study - N_society, V_suggest - J_other, V_support - N_family, V_take - J_many, V_take - N_part, V_tend - J_large, V_tend - J_small, V_tend - N_society, V_think - J_many, V_think - J_most, V_understand - N_world, V_use - J_many, V_use - J_other, V_use - N_group, V_vary - N_country, V_vary - N_group, V_vary - N_population, V_vary - N_society, V_vary - N_state, V_work - N_economy, V_work - V_live
Collocational Dimension 7: Reporting Research (658 collocations)
J_able - V_explain, J_anthropological - N_research, J_average - V_show, J_basic - N_research, J_certain - V_explain, J_clinical - N_research, J_consistent - J_empirical, J_consistent - J_previous, J_consistent - N_finding, J_consistent - N_research, J_current - J_future, J_different - V_base, J_different - V_come, J_different - V_show, J_early - N_research, J_early - V_suggest, J_empirical - J_consistent, J_empirical - J_recent, J_empirical - N_finding, J_empirical - N_literature, J_empirical - N_research, J_empirical - V_support, J_environmental - J_genetic, J_ethnographic - N_research, J_existing - N_literature, J_experimental - N_research, J_extensive - N_literature, J_few - V_examine, J_first - V_come, J_further - N_research, J_future - N_research, J_future - V_examine, J_genetic - V_suggest, J_great - V_show, J_high - V_indicate, J_high - V_show, J_human - J_early, J_human - J_genetic, J_human - J_recent, J_human - V_explain, J_important - J_future, J_important - N_finding, J_important - N_research, J_individual - V_focus, J_interesting - J_future, J_interesting - N_finding, J_key - N_finding, J_large - N_literature, J_late - J_early, J_likely - V_report, J_main - N_finding, J_more - J_recent, J_more - N_research, J_more - V_report, J_most - N_research, J_much - N_research, J_negative - V_indicate, J_new - J_genetic, J_new - N_research, J_new - V_come, J_next - V_describe, J_next - V_examine, J_nineteenth - J_early, J_online - V_present, J_organizational - N_research, J_other - N_research, J_other - V_show, J_other - V_suggest, J_particular - V_focus, J_past - J_future, J_past - N_research, J_positive - N_finding, J_positive - V_indicate, J_positive - V_show, J_present - J_future, J_previous - J_consistent, J_previous - N_finding, J_previous - N_literature, J_previous - N_research, J_previous - V_describe, J_previous - V_examine, J_previous - V_suggest, J_prior - J_consistent, J_prior - N_research, J_prior - V_suggest, J_psychological - N_literature, J_psychological - N_research, J_recent - J_empirical, J_recent - N_finding, J_recent - N_literature, J_recent - N_research, J_recent - V_indicate, J_recent - V_show, J_recent - V_suggest, J_related - N_literature, J_relevant - N_finding, J_relevant - N_literature, J_robust - N_finding, J_same - V_show, J_scientific - N_research, J_several - V_show, J_significant - V_indicate, J_significant - V_report, J_significant - V_show, J_similar - N_finding, J_similar - V_report, J_similar - V_show, J_single - V_focus, J_social - N_research, J_social - V_base, J_social - V_examine, J_social - V_explain, J_social - V_focus, J_specific - V_focus, J_subsequent - N_research, J_such - N_research, J_theoretical - J_empirical, J_theoretical - N_finding, J_theoretical - V_present, J_twentieth - J_early, J_various - V_examine, N_activity - V_show, N_advance -

J_recent, N_age - J_early, N_age - V_base, N_analysis - J_empirical, N_analysis - V_base, N_analysis - V_examine, N_analysis - V_focus, N_analysis - V_indicate, N_analysis - V_present, N_analysis - V_show, N_analysis - V_suggest, N_answer - V_come, N_approach - V_describe, N_approach - V_focus, N_approach - V_suggest, N_area - N_research, N_argument - V_support, N_article - V_describe, N_article - V_examine, N_article - V_focus, N_article - V_present, N_article - V_report, N_aspect - V_focus, N_association - V_report, N_association - V_suggest, N_assumption - V_base, N_attempt - V_explain, N_attention - V_focus, N_author - V_suggest, N_basis - J_genetic, N_behavior - V_explain, N_belief - V_base, N_body - N_literature, N_body - N_research, N_category - V_base, N_century - J_early, N_challenge - V_present, N_change - J_genetic, N_change - V_show, N_chapter - J_early, N_chapter - J_previous, N_chapter - V_describe, N_chapter - V_examine, N_chapter - V_focus, N_chapter - V_present, N_characteristic - V_base, N_characteristic - V_describe, N_child - V_show, N_childhood - J_early, N_claim - V_support, N_coefficient - V_indicate, N_colleague - V_suggest, N_column - V_report, N_column - V_show, N_component - J_genetic, N_concept - V_base, N_conclusion - N_finding, N_conclusion - V_come, N_conclusion - V_support, N_contact - V_come, N_contribution - N_literature, N_correlation - V_report, N_criterion - V_base, N_culture - V_base, N_date - J_early, N_datum - J_consistent, N_datum - J_empirical, N_datum - J_recent, N_datum - V_base, N_datum - V_come, N_datum - V_describe, N_datum - V_examine, N_datum - V_indicate, N_datum - V_present, N_datum - V_report, N_datum - V_show, N_datum - V_suggest, N_datum - V_support, N_day - J_early, N_decade - J_recent, N_decision - V_base, N_design - N_research, N_detail - V_describe, N_detail - V_examine, N_development - J_early, N_development - J_recent, N_development - N_research, N_difference - J_genetic, N_difference - V_base, N_difference - V_explain, N_difference - V_show, N_difference - V_suggest, N_direction - J_future, N_direction - N_research, N_discussion - J_early, N_discussion - J_previous, N_discussion - V_focus, N_diversity - J_genetic, N_economics - N_literature, N_economy - V_base, N_effect - N_research, N_effect - V_examine, N_effect - V_report, N_effect - V_show, N_effect - V_suggest, N_effort - V_focus, N_effort - V_support, N_end - V_come, N_equation - V_describe, N_equation - V_show, N_era - J_early, N_error - V_report, N_estimate - J_consistent, N_estimate - V_base, N_estimate - V_indicate, N_estimate - V_present, N_estimate - V_report, N_estimate - V_suggest, N_event - J_future, N_event - J_recent, N_event - V_describe, N_evidence - J_consistent, N_evidence - J_early, N_evidence - J_empirical, N_evidence - J_genetic, N_evidence - J_recent, N_evidence - N_finding, N_evidence - V_base, N_evidence - V_come, N_evidence - V_indicate, N_evidence - V_present, N_evidence - V_show, N_evidence - V_suggest, N_evidence - V_support, N_example - J_previous, N_example - J_recent, N_example - V_show, N_expectation - J_future, N_experience - J_early, N_experience - J_previous, N_experience - J_recent, N_experience - V_base, N_experience - V_describe, N_experiment - V_describe, N_experiment - V_present, N_experiment - V_show, N_extent - V_examine, N_fact - J_consistent, N_fact - V_base, N_fact - V_explain, N_factor - J_genetic, N_factor - V_explain, N_factor - V_focus, N_family - V_support, N_field - N_research, N_figure - V_show, N_finding - J_consistent, N_finding - J_empirical, N_finding - J_previous, N_finding - J_recent, N_finding - N_literature, N_finding - N_research, N_finding - V_base, N_finding - V_explain, N_finding - V_indicate, N_finding - V_present, N_finding - V_report, N_finding - V_show, N_finding - V_suggest, N_finding - V_support, N_form - J_early, N_form - V_come, N_gender - V_base, N_gene - J_genetic, N_generation - J_future, N_generation - J_previous, N_graph - V_show, N_group - V_base, N_group - V_show, N_history - J_recent, N_human - J_early, N_human - V_suggest, N_hypothesis - J_consistent, N_hypothesis - N_research, N_hypothesis - V_explain, N_hypothesis - V_suggest, N_hypothesis - V_support, N_idea - J_consistent, N_idea - N_research, N_idea - V_base, N_idea - V_come, N_idea - V_support, N_impact - V_examine, N_implication - N_finding, N_implication - N_research, N_importance - N_finding, N_income - V_report, N_increase - V_show, N_inference - V_base, N_influence - J_genetic, N_information - V_base, N_information - V_present, N_interaction - V_examine, N_interpretation - J_consistent, N_interpretation - V_support, N_investigation - J_empirical, N_issue - N_research, N_issue - V_examine, N_issue - V_focus, N_judgment - V_base, N_knowledge - V_base, N_life - J_early, N_light - N_finding, N_line - N_research, N_line - V_indicate, N_line - V_show, N_line - V_suggest, N_literature - J_empirical, N_literature - J_previous, N_literature - J_recent, N_literature - N_finding, N_literature - N_research, N_literature - V_examine, N_literature - V_focus, N_literature - V_suggest, N_measure - V_base, N_mechanism - J_genetic, N_medium - V_report, N_method - N_research, N_method - V_base, N_method - V_describe, N_model - J_consistent, N_model - V_base, N_model - V_describe, N_model - V_explain, N_model

- V_present, N_model - V_show, N_model - V_suggest, N_notion - J_consistent, N_notion - V_support, N_number - V_indicate, N_number - V_report, N_number - V_show, N_observation - V_base, N_opportunity - V_present, N_origin - V_explain, N_outcome - J_future, N_outcome - V_focus, N_outcome - V_report, N_panel - V_show, N_paper - J_recent, N_paper - N_literature, N_paper - V_focus, N_paper - V_present, N_part - J_early, N_participant - N_research, N_participant - V_indicate, N_participant - V_present, N_participant - V_report, N_participant - V_show, N_past - J_recent, N_patient - V_show, N_pattern - J_consistent, N_pattern - V_examine, N_pattern - V_explain, N_pattern - V_show, N_pattern - V_suggest, N_people - V_come, N_people - V_show, N_people - V_suggest, N_percent - V_report, N_performance - V_base, N_performance - V_explain, N_period - J_early, N_period - J_previous, N_phenomenon - J_recent, N_phenomenon - V_explain, N_picture - V_present, N_play - V_come, N_power - V_come, N_practice - N_research, N_prediction - J_consistent, N_prediction - V_base, N_presence - V_indicate, N_price - J_future, N_principle - V_base, N_problem - V_present, N_procedure - V_describe, N_process - N_research, N_process - V_describe, N_project - N_research, N_question - J_empirical, N_question - N_research, N_question - V_examine, N_question - V_focus, N_race - V_base, N_rate - J_future, N_rate - V_show, N_regression - V_present, N_relationship - N_research, N_relationship - V_describe, N_relationship - V_examine, N_relationship - V_show, N_report - J_recent, N_research - J_consistent, N_research - J_early, N_research - J_empirical, N_research - J_future, N_research - J_previous, N_research - J_recent, N_research - N_finding, N_research - N_literature, N_research - V_base, N_research - V_describe, N_research - V_examine, N_research - V_focus, N_research - V_indicate, N_research - V_show, N_research - V_suggest, N_research - V_support, N_researcher - V_examine, N_researcher - V_focus, N_researcher - V_suggest, N_resource - V_support, N_respondent - V_indicate, N_respondent - V_report, N_response - V_show, N_result - J_consistent, N_result - J_empirical, N_result - N_research, N_result - V_base, N_result - V_explain, N_result - V_indicate, N_result - V_present, N_result - V_report, N_result - V_show, N_result - V_suggest, N_result - V_support, N_return - J_future, N_review - J_recent, N_review - N_literature, N_review - N_research, N_review - V_focus, N_role - V_examine, N_role - V_focus, N_sample - V_base, N_science - N_literature, N_science - N_research, N_score - V_indicate, N_section - J_previous, N_section - V_describe, N_section - V_examine, N_section - V_focus, N_section - V_present, N_section - V_show, N_series - V_present, N_setting - N_research, N_sign - V_show, N_site - V_come, N_situation - V_describe, N_size - V_base, N_source - V_come, N_stage - J_early, N_statistic - V_present, N_statistic - V_report, N_status - V_base, N_stimulus - V_present, N_strategy - J_empirical, N_strategy - N_research, N_student - V_report, N_study - J_early, N_study - J_empirical, N_study - J_future, N_study - J_genetic, N_study - J_previous, N_study - J_recent, N_study - N_finding, N_study - N_research, N_study - V_base, N_study - V_come, N_study - V_describe, N_study - V_examine, N_study - V_focus, N_study - V_indicate, N_study - V_report, N_study - V_show, N_study - V_suggest, N_study - V_support, N_subject - N_research, N_support - J_empirical, N_survey - J_recent, N_survey - N_research, N_system - V_base, N_table - V_present, N_table - V_report, N_table - V_show, N_task - V_focus, N_team - N_research, N_term - V_describe, N_term - V_explain, N_test - J_empirical, N_test - V_base, N_theory - J_consistent, N_theory - N_research, N_theory - V_base, N_theory - V_describe, N_theory - V_explain, N_theory - V_focus, N_theory - V_suggest, N_theory - V_support, N_topic - N_research, N_trait - J_genetic, N_trend - J_recent, N_trial - V_present, N_value - J_future, N_value - V_base, N_variable - V_indicate, N_variance - V_explain, N_variation - J_genetic, N_variation - V_explain, N_view - J_consistent, N_view - V_support, N_way - V_describe, N_woman - V_report, N_word - V_come, N_word - V_present, N_work - J_early, N_work - J_empirical, N_work - J_future, N_work - J_recent, N_work - V_focus, N_work - V_show, N_work - V_suggest, N_year - J_early, N_year - J_previous, N_year - J_recent, N_year - V_come, V_account - N_finding, V_address - N_literature, V_address - N_research, V_appear - J_early, V_ask - V_indicate, V_ask - V_report, V_attempt - V_explain, V_base - N_finding, V_base - N_research, V_begin - J_early, V_begin - N_research, V_concern - N_literature, V_conduct - N_research, V_confirm - N_finding, V_contribute - N_literature, V_demonstrate - J_previous, V_demonstrate - J_recent, V_demonstrate - N_finding, V_demonstrate - N_literature, V_demonstrate - N_research, V_depend - J_future, V_describe - J_previous, V_describe - N_research, V_develop - N_research, V_discuss - J_previous, V_discuss - N_finding, V_draw - N_literature, V_emerge - N_literature, V_emphasize - N_literature, V_examine - J_future, V_examine - J_previous, V_examine - N_literature, V_examine - N_research, V_exist - N_literature, V_expect - J_future, V_expect - V_base, V_explain - N_finding, V_explore - J_future, V_explore - N_research, V_extend -

N_finding, V_feel - V_report, V_find - J_early, V_find - N_research, V_focus - N_literature, V_focus - N_research, V_generalize - N_finding, V_go - V_come, V_grow - N_literature, V_help - V_explain, V_highlight - N_finding, V_identify - N_research, V_include - N_research, V_increase - V_show, V_indicate - J_recent, V_indicate - N_finding, V_indicate - N_research, V_interpret - N_finding, V_investigate - J_future, V_investigate - N_research, V_involve - N_research, V_know - J_early, V_know - V_come, V_lead - N_research, V_let - V_examine, V_need - N_research, V_need - V_support, V_predict - J_future, V_present - N_finding, V_provide - J_empirical, V_provide - N_finding, V_publish - N_research, V_regard - N_finding, V_relate - N_literature, V_report - N_finding, V_reveal - N_finding, V_reveal - N_research, V_review - N_literature, V_review - N_research, V_see - J_early, V_seek - V_explain, V_show - J_recent, V_show - N_finding, V_show - N_research, V_suggest - J_early, V_suggest - J_genetic, V_suggest - J_previous, V_suggest - J_recent, V_suggest - N_finding, V_suggest - N_literature, V_suggest - N_research, V_summarize - N_finding, V_support - J_empirical, V_support - N_finding, V_support - N_research, V_tend - V_focus, V_try - V_explain, V_understand - V_come, V_use - N_research, V_use - V_describe
Collocational Dimension 8: Individuals in Society (849 collocations)
J_additional - V_receive, J_american - N_man, J_american - N_student, J_american - N_woman, J_average - N_age, J_black - N_man, J_black - N_woman, J_certain - N_people, J_complex - J_more, J_cultural - N_gender, J_daily - N_life, J_different - N_age, J_different - N_individual, J_different - N_people, J_different - V_compare, J_different - V_find, J_difficult - V_find, J_early - N_age, J_early - N_life, J_early - N_work, J_early - V_find, J_efficient - J_more, J_empirical - N_work, J_entire - N_life, J_everyday - N_life, J_female - N_student, J_few - J_more, J_few - N_individual, J_few - N_people, J_few - V_work, J_first - N_age, J_formal - N_education, J_future - N_work, J_gay - N_man, J_good - J_more, J_good - N_life, J_good - N_people, J_good - V_find, J_good - V_work, J_great - N_man, J_hard - N_work, J_hard - V_find, J_hard - V_work, J_healthy - N_life, J_high - J_more, J_high - N_education, J_high - N_individual, J_high - N_people, J_high - N_school, J_high - N_student, J_high - V_receive, J_high - V_tend, J_human - N_life, J_indigenous - N_people, J_large - J_more, J_large - N_people, J_large - V_find, J_large - V_tend, J_late - N_life, J_least - N_individual, J_less - J_more, J_less - J_old, J_less - V_receive, J_less - V_work, J_likely - N_child, J_likely - N_individual, J_likely - N_man, J_likely - N_people, J_likely - N_woman, J_likely - V_receive, J_little - V_receive, J_local - N_people, J_local - N_school, J_long - N_life, J_long - V_work, J_low - J_poor, J_low - V_tend, J_male - J_young, J_male - N_gender, J_male - N_student, J_many - J_more, J_many - N_child, J_many - N_individual, J_many - N_life, J_many - N_man, J_many - N_people, J_many - N_student, J_many - N_woman, J_many - V_find, J_married - N_child, J_married - N_man, J_married - N_woman, J_mean - N_age, J_medical - N_school, J_mental - N_people, J_middle - N_age, J_middle - N_school, J_modern - N_people, J_more - N_child, J_more - N_individual, J_more - N_man, J_more - N_people, J_more - N_student, J_more - N_woman, J_more - N_work, J_more - V_find, J_more - V_receive, J_more - V_tend, J_more - V_work, J_most - N_man, J_most - N_people, J_most - N_student, J_most - N_woman, J_most - V_find, J_much - N_work, J_native - N_people, J_new - J_old, J_new - V_find, J_old - J_young, J_old - N_age, J_old - N_child, J_old - N_individual, J_old - N_man, J_old - N_people, J_old - N_woman, J_old - V_tend, J_other - J_more, J_other - N_child, J_other - N_individual, J_other - N_life, J_other - N_people, J_other - N_work, J_other - V_compare, J_other - V_find, J_own - N_child, J_own - N_home, J_own - N_life, J_particular - N_individual, J_personal - N_life, J_poor - N_child, J_poor - N_people, J_poor - N_woman, J_poor - V_work, J_primary - N_school, J_prior - N_work, J_private - N_school, J_public - N_education, J_public - N_school, J_racial - N_gender, J_real - N_life, J_recent - J_more, J_recent - N_work, J_rich - J_poor, J_rural - J_poor, J_same - N_age, J_same - N_individual, J_same - N_people, J_same - V_find, J_secondary - N_school, J_several - V_find, J_sexual - N_man, J_sexual - N_woman, J_significant - V_find, J_similar - V_find, J_similar - V_tend, J_single - N_individual, J_single - N_man, J_single - N_woman, J_small - N_individual, J_small - V_find, J_small - V_tend, J_social - N_individual, J_social - N_life, J_social - N_people, J_such - J_more, J_such - N_gender, J_such - N_people, J_such - N_work, J_such - V_find, J_urban - N_life, J_wealthy - J_poor, J_white - N_man, J_white - N_woman, J_young - J_old, J_young - N_age, J_young - N_child, J_young - N_individual, J_young - N_man, J_young - N_people, J_young - N_woman, N_ability - N_child, N_ability - N_individual, N_abuse - N_child, N_access - N_education, N_activity - N_work, N_adult - J_likely, N_adult - J_old, N_adult - J_young, N_adult - N_age, N_adult - N_child, N_adult - N_man, N_age - J_old, N_age - J_young, N_age - N_child, N_age -

N_education, N_age - N_gender, N_age - N_individual, N_age - N_people, N_age - N_woman, N_amount - V_receive, N_animal - N_life, N_aspect - N_life, N_association - V_find, N_attention - J_more, N_attention - V_receive, N_attitude - N_people, N_behavior - N_child, N_behavior - N_individual, N_behavior - N_people, N_behavior - N_work, N_belief - N_people, N_benefit - V_receive, N_billion - J_more, N_birth - N_child, N_birth - N_life, N_birth - N_woman, N_boy - J_young, N_boy - N_man, N_care - N_child, N_care - N_education, N_care - V_receive, N_category - N_age, N_category - N_gender, N_century - J_more, N_chance - N_life, N_characteristic - N_individual, N_child - J_likely, N_child - J_more, N_child - J_old, N_child - J_poor, N_child - J_young, N_child - N_age, N_child - N_education, N_child - N_home, N_child - N_man, N_child - N_school, N_child - N_woman, N_child - V_compare, N_child - V_find, N_child - V_receive, N_child - V_tend, N_class - N_age, N_class - N_gender, N_class - N_school, N_class - N_student, N_class - V_work, N_colleague - V_find, N_college - N_education, N_college - N_school, N_college - N_student, N_color - N_woman, N_community - J_poor, N_community - N_home, N_community - N_individual, N_community - N_life, N_community - N_people, N_company - J_more, N_company - V_work, N_compensation - V_receive, N_condition - J_poor, N_condition - V_compare, N_condition - V_work, N_control - V_compare, N_cost - V_compare, N_country - J_more, N_country - J_poor, N_country - N_home, N_country - N_people, N_course - N_life, N_course - N_student, N_culture - N_people, N_cycle - N_life, N_datum - J_more, N_day - N_school, N_day - V_compare, N_day - V_work, N_death - N_age, N_death - N_life, N_detail - J_more, N_development - N_child, N_difference - N_age, N_difference - N_gender, N_difference - V_find, N_economy - V_work, N_education - N_age, N_education - N_child, N_education - N_school, N_education - V_receive, N_effect - N_gender, N_effect - V_compare, N_effect - V_find, N_effort - N_work, N_employee - V_work, N_employment - N_education, N_environment - N_work, N_equality - N_gender, N_estimate - V_compare, N_ethnicity - N_age, N_ethnicity - N_gender, N_event - N_life, N_evidence - V_find, N_example - N_people, N_example - V_find, N_experience - N_life, N_experience - N_woman, N_experience - N_work, N_fact - N_people, N_family - J_poor, N_family - N_child, N_family - N_education, N_family - N_home, N_family - N_individual, N_family - N_life, N_family - N_school, N_family - N_woman, N_family - N_work, N_father - N_child, N_feedback - V_receive, N_female - J_young, N_field - N_work, N_firm - J_more, N_firm - V_receive, N_food - J_more, N_food - V_find, N_force - N_work, N_force - V_work, N_form - N_life, N_gap - N_gender, N_gender - N_age, N_gender - N_man, N_gender - N_woman, N_generation - J_young, N_girl - J_young, N_girl - N_woman, N_grade - N_school, N_grade - N_student, N_group - J_more, N_group - N_age, N_group - N_individual, N_group - N_people, N_group - N_student, N_group - N_work, N_group - V_compare, N_half - J_more, N_health - J_poor, N_health - N_child, N_health - N_education, N_health - N_life, N_health - N_woman, N_health - V_receive, N_help - V_receive, N_history - N_life, N_home - N_child, N_home - N_school, N_home - N_woman, N_home - N_work, N_home - V_work, N_hospital - N_school, N_hour - J_more, N_hour - N_work, N_hour - V_work, N_household - N_child, N_human - V_compare, N_idea - N_people, N_identity - N_gender, N_income - J_more, N_income - N_education, N_income - N_people, N_income - V_receive, N_increase - N_age, N_individual - J_likely, N_individual - J_more, N_individual - J_old, N_individual - J_young, N_individual - N_age, N_individual - N_life, N_individual - V_find, N_inequality - N_gender, N_infant - N_child, N_information - J_more, N_information - V_receive, N_instruction - V_receive, N_insurance - N_life, N_job - N_education, N_job - N_work, N_job - V_find, N_job - V_receive, N_job - V_work, N_kind - N_people, N_kind - N_work, N_labor - N_child, N_labor - N_woman, N_land - N_home, N_law - N_school, N_level - N_education, N_life - N_individual, N_life - N_people, N_life - N_woman, N_life - N_work, N_line - N_gender, N_loan - N_student, N_male - J_likely, N_male - J_young, N_man - J_likely, N_man - J_more, N_man - J_old, N_man - J_young, N_man - N_child, N_man - N_gender, N_man - N_woman, N_man - N_work, N_man - V_compare, N_man - V_find, N_man - V_tend, N_man - V_work, N_marriage - N_man, N_marriage - N_woman, N_memory - V_work, N_message - V_receive, N_million - N_people, N_minority - N_woman, N_model - V_compare, N_money - J_more, N_money - N_people, N_money - V_receive, N_month - N_age, N_mortgage - N_home, N_mother - J_young, N_mother - N_child, N_neighborhood - J_poor, N_neighborhood - N_school, N_number - N_child, N_number - N_individual, N_number - N_people, N_number - N_woman, N_office - N_home, N_one - J_old, N_organization - N_individual, N_other - J_likely, N_other - J_more, N_other - N_people, N_outcome - J_poor, N_outcome - V_compare, N_parent - N_child, N_parent - N_home, N_parent - N_student, N_part - N_life, N_participant - V_receive, N_participation - N_woman, N_pattern - V_find, N_pay - V_receive, N_payment - V_receive, N_people - J_likely,

N_people - J_more, N_people - J_old, N_people - J_poor, N_people - J_young, N_people - N_age, N_people - N_life, N_people - V_find, N_people - V_tend, N_people - V_work, N_percent - J_more, N_percent - N_man, N_percent - N_woman, N_percent - V_compare, N_percent - V_receive, N_performance - J_poor, N_performance - N_student, N_performance - V_compare, N_person - N_life, N_person - V_receive, N_place - N_home, N_place - N_people, N_place - N_work, N_population - N_age, N_population - N_individual, N_poverty - J_poor, N_practice - N_work, N_problem - V_work, N_program - N_education, N_program - N_school, N_quality - J_poor, N_quality - N_life, N_quality - N_school, N_race - N_age, N_race - N_gender, N_range - N_age, N_rate - V_find, N_region - V_find, N_relation - N_gender, N_relationship - N_age, N_relationship - N_individual, N_relationship - V_find, N_research - J_more, N_researcher - V_find, N_researcher - V_work, N_resource - J_more, N_result - V_compare, N_result - V_find, N_reward - V_receive, N_right - N_woman, N_role - N_gender, N_role - N_life, N_role - N_man, N_role - N_woman, N_role - N_work, N_sample - N_age, N_sample - N_individual, N_sample - N_student, N_school - N_child, N_school - N_education, N_school - N_home, N_school - N_student, N_scientist - V_work, N_score - N_student, N_service - V_receive, N_sex - N_age, N_sex - N_gender, N_sex - N_individual, N_sex - N_man, N_share - V_receive, N_site - V_find, N_size - V_compare, N_skill - N_child, N_skill - N_education, N_society - N_individual, N_society - N_man, N_society - N_people, N_society - N_woman, N_society - V_find, N_society - V_tend, N_stage - N_life, N_status - N_age, N_status - N_gender, N_status - N_woman, N_student - J_more, N_student - N_school, N_student - V_find, N_student - V_receive, N_student - V_work, N_study - J_more, N_study - N_child, N_study - N_student, N_study - N_woman, N_study - V_compare, N_study - V_find, N_support - V_find, N_support - V_receive, N_system - N_school, N_system - V_work, N_teacher - N_school, N_teacher - N_student, N_team - N_work, N_team - V_work, N_test - N_student, N_thing - N_people, N_time - J_likely, N_time - J_more, N_time - N_people, N_time - V_work, N_tool - V_find, N_training - N_education, N_training - V_receive, N_treatment - V_receive, N_university - N_student, N_value - J_more, N_value - N_life, N_value - V_find, N_variable - J_more, N_wage - N_woman, N_wage - V_receive, N_wage - V_work, N_way - N_life, N_way - N_people, N_way - V_find, N_way - V_work, N_week - V_work, N_wife - N_child, N_woman - J_likely, N_woman - J_more, N_woman - J_old, N_woman - J_poor, N_woman - J_young, N_woman - N_age, N_woman - N_child, N_woman - N_gender, N_woman - N_home, N_woman - N_life, N_woman - N_man, N_woman - N_work, N_woman - V_compare, N_woman - V_find, N_woman - V_tend, N_woman - V_work, N_work - J_more, N_work - N_home, N_work - N_life, N_work - N_man, N_work - N_woman, N_work - V_find, N_worker - J_more, N_worker - J_old, N_worker - J_young, N_worker - N_work, N_worker - V_receive, N_worker - V_work, N_world - J_poor, N_world - N_people, N_year - J_more, N_year - J_old, N_year - N_age, N_year - N_education, N_year - N_life, N_year - N_school, V_add - J_more, V_adopt - N_child, V_affect - N_individual, V_affect - N_life, V_affect - N_people, V_allow - J_more, V_allow - N_individual, V_allow - N_people, V_allow - V_compare, V_ask - N_child, V_ask - N_people, V_ask - N_student, V_attend - N_child, V_attend - N_school, V_attend - N_student, V_base - N_age, V_base - N_gender, V_bear - N_child, V_bear - N_individual, V_become - J_likely, V_become - J_more, V_become - N_individual, V_become - N_people, V_become - N_woman, V_begin - N_age, V_believe - N_people, V_bring - N_home, V_build - N_work, V_buy - J_more, V_care - N_child, V_carry - N_individual, V_change - N_people, V_come - N_home, V_come - N_people, V_compare - N_child, V_compare - N_man, V_compare - N_woman, V_concentrate - V_tend, V_consider - N_people, V_consider - N_work, V_create - J_more, V_create - N_people, V_demonstrate - N_work, V_develop - N_child, V_die - J_likely, V_die - N_age, V_differ - N_gender, V_differ - N_individual, V_discuss - J_more, V_drop - N_school, V_earn - J_more, V_engage - J_likely, V_engage - N_child, V_engage - N_people, V_expect - V_find, V_expect - V_receive, V_experience - J_likely, V_experience - N_child, V_experience - N_individual, V_experience - N_people, V_fall - V_tend, V_favor - V_tend, V_feel - N_individual, V_feel - N_people, V_find - J_more, V_find - N_child, V_find - N_individual, V_find - N_man, V_find - N_people, V_find - N_student, V_find - N_woman, V_find - N_work, V_focus - N_work, V_focus - V_tend, V_get - J_more, V_get - J_old, V_get - N_people, V_give - J_more, V_give - N_child, V_give - N_individual, V_give - N_people, V_give - N_student, V_give - N_woman, V_go - N_home, V_go - N_school, V_go - V_work, V_grow - J_old, V_grow - N_child, V_help - J_poor, V_help - N_child, V_help - N_people, V_help - N_student, V_help - N_work, V_identify - J_likely, V_identify - N_individual, V_improve - N_life, V_include - N_child, V_include - N_people, V_include - N_work, V_increase - N_age, V_increase - V_tend, V_involve - N_individual, V_involve - N_work, V_kill - N_man, V_kill - N_people, V_know -

N_people, V_lead - J_more, V_lead - N_life, V_learn - J_more, V_learn - J_young, V_learn - N_child, V_learn - N_individual, V_learn - N_people, V_learn - N_student, V_leave - N_child, V_leave - N_home, V_leave - N_woman, V_live - N_age, V_live - N_child, V_live - N_home, V_live - N_individual, V_live - N_life, V_live - N_man, V_live - N_people, V_live - N_woman, V_live - V_work, V_make - J_likely, V_make - J_more, V_make - N_individual, V_make - N_life, V_make - N_people, V_make - N_woman, V_marry - N_man, V_marry - N_woman, V_mean - J_more, V_mean - N_people, V_move - N_people, V_need - J_more, V_need - N_work, V_need - V_work, V_occur - J_likely, V_own - N_home, V_participate - N_student, V_pay - J_more, V_pay - N_people, V_pay - N_work, V_perform - N_work, V_play - N_child, V_possess - N_individual, V_produce - J_more, V_provide - J_more, V_provide - N_people, V_raise - N_child, V_reach - N_age, V_read - N_student, V_receive - J_likely, V_receive - J_more, V_receive - N_child, V_receive - N_education, V_receive - N_student, V_rely - V_tend, V_replace - J_old, V_report - J_likely, V_report - J_more, V_report - N_student, V_report - N_woman, V_represent - N_individual, V_require - J_more, V_require - N_student, V_respond - N_individual, V_return - N_home, V_save - N_life, V_say - N_people, V_say - N_woman, V_see - J_more, V_see - N_people, V_seek - N_people, V_seem - J_likely, V_sell - J_more, V_send - N_child, V_shape - N_life, V_share - N_individual, V_share - N_people, V_show - N_child, V_show - N_people, V_show - N_work, V_spend - J_more, V_stay - N_home, V_study - N_people, V_study - N_student, V_suggest - N_people, V_suggest - N_work, V_survive - J_likely, V_take - J_more, V_take - N_people, V_take - N_student, V_teach - N_child, V_teach - N_school, V_teach - N_student, V_tend - J_more, V_tend - J_old, V_tend - N_child, V_tend - N_man, V_tend - N_people, V_tend - N_woman, V_think - N_people, V_think - V_tend, V_train - N_student, V_try - V_find, V_understand - N_people, V_use - J_more, V_use - N_people, V_use - V_find, V_vary - N_individual, V_want - N_people, V_want - V_work, V_watch - N_child, V_work - J_more, V_work - J_poor, V_work - N_home, V_work - N_man, V_work - N_people, V_work - N_student, V_work - N_woman
Collocational Dimension 9: Gaining Knowledge and Developing Concepts (748 collocations)
J_accurate - N_information, J_additional - N_information, J_additional - V_need, J_additional - V_produce, J_additional - V_require, J_alternative - N_use, J_available - N_information, J_available - V_make, J_bad - V_get, J_clear - V_make, J_common - J_new, J_common - N_use, J_common - V_share, J_complex - V_require, J_considerable - V_require, J_cultural - V_share, J_detailed - N_information, J_different - N_information, J_different - V_give, J_different - V_make, J_different - V_produce, J_different - V_take, J_different - V_use, J_difficult - V_make, J_easy - V_make, J_efficient - N_use, J_enough - V_produce, J_extensive - N_use, J_few - V_produce, J_full - N_information, J_further - V_need, J_further - V_require, J_genetic - J_new, J_good - V_get, J_good - V_give, J_good - V_make, J_good - V_produce, J_good - V_require, J_great - V_give, J_great - V_make, J_hard - V_get, J_high - V_produce, J_important - N_information, J_important - V_make, J_large - V_get, J_large - V_make, J_large - V_produce, J_large - V_require, J_less - V_produce, J_likely - V_make, J_limited - N_information, J_long - V_take, J_many - J_new, J_many - V_make, J_many - V_produce, J_many - V_share, J_many - V_take, J_many - V_use, J_married - V_get, J_more - N_information, J_more - V_allow, J_more - V_get, J_more - V_give, J_more - V_learn, J_more - V_make, J_more - V_need, J_more - V_produce, J_more - V_require, J_more - V_take, J_more - V_use, J_much - N_information, J_multiple - V_allow, J_natural - V_produce, J_new - J_common, J_new - N_information, J_new - V_allow, J_new - V_develop, J_new - V_give, J_new - V_learn, J_new - V_make, J_new - V_need, J_new - V_produce, J_new - V_require, J_new - V_take, J_old - J_new, J_old - V_get, J_online - J_available, J_online - N_information, J_other - N_information, J_other - V_make, J_other - V_use, J_own - V_develop, J_own - V_make, J_own - V_produce, J_possible - V_make, J_private - N_information, J_relevant - N_information, J_rich - V_get, J_same - V_apply, J_same - V_give, J_same - V_make, J_same - V_produce, J_same - V_share, J_same - V_use, J_sensory - N_information, J_similar - V_apply, J_similar - V_produce, J_similar - V_share, J_similar - V_use, J_simple - V_use, J_social - J_new, J_special - V_require, J_specific - N_information, J_standard - V_use, J_statistical - V_use, J_successful - V_require, J_such - J_new, J_such - N_information, J_such - N_use, J_such - V_make, J_such - V_use, J_sure - V_make, J_useful - N_information, J_valuable - N_information, J_visual - N_information, J_widespread - N_use, J_young - V_learn, N_ability - V_develop, N_ability - V_learn, N_ability - V_produce, N_account - V_take, N_action - V_take, N_adjustment - V_make, N_advantage - V_give, N_advantage - V_take, N_alcohol - N_use, N_alternative - J_available, N_amount - J_available, N_amount - N_information, N_amount - V_give, N_amount - V_need,

N_amount - V_produce, N_amount - V_require, N_analysis - V_apply, N_analysis - V_require, N_analysis - V_use, N_ancestor - J_common, N_answer - V_give, N_approach - J_common, N_approach - J_new, N_approach - V_allow, N_approach - V_apply, N_approach - V_develop, N_approach - V_require, N_approach - V_take, N_approach - V_use, N_argument - V_apply, N_assumption - V_make, N_assumption - V_require, N_attention - V_require, N_author - J_available, N_bank - V_make, N_behavior - V_learn, N_behavior - V_produce, N_belief - J_common, N_belief - V_share, N_birth - V_give, N_business - J_new, N_business - V_make, N_capital - J_new, N_car - J_new, N_care - V_take, N_case - V_make, N_cell - V_produce, N_challenge - J_new, N_chance - V_get, N_change - V_make, N_change - V_produce, N_characteristic - V_share, N_child - V_give, N_child - V_learn, N_choice - V_give, N_choice - V_make, N_claim - V_make, N_communication - N_information, N_company - J_new, N_company - V_allow, N_company - V_make, N_company - V_use, N_comparison - V_allow, N_comparison - V_make, N_concept - V_apply, N_concept - V_develop, N_condition - V_give, N_condition - V_learn, N_context - V_learn, N_contribution - V_make, N_cost - J_new, N_cost - V_give, N_cost - V_produce, N_country - V_develop, N_country - V_produce, N_course - V_take, N_credit - V_get, N_criterion - V_apply, N_culture - V_develop, N_culture - V_learn, N_culture - V_share, N_datum - J_available, N_datum - N_information, N_datum - N_use, N_datum - V_apply, N_datum - V_give, N_datum - V_use, N_day - V_give, N_day - V_take, N_debt - V_use, N_decision - V_make, N_decision - V_require, N_descent - J_common, N_design - V_use, N_development - J_new, N_difference - V_make, N_difficulty - V_learn, N_discovery - J_new, N_disease - V_develop, N_distinction - V_make, N_drug - N_use, N_economy - V_develop, N_economy - V_produce, N_effect - V_produce, N_effect - V_use, N_effort - V_make, N_effort - V_require, N_element - J_common, N_energy - V_require, N_entry - J_new, N_environment - J_new, N_environment - V_learn, N_equation - V_give, N_equation - V_use, N_error - V_make, N_estimate - V_give, N_estimate - V_produce, N_estimate - V_use, N_evidence - J_available, N_evidence - J_new, N_example - V_give, N_example - V_take, N_example - V_use, N_experience - V_learn, N_experience - V_share, N_experiment - V_use, N_factor - J_common, N_feature - J_common, N_feature - V_share, N_field - V_apply, N_firm - J_new, N_firm - V_allow, N_firm - V_give, N_firm - V_make, N_firm - V_need, N_firm - V_produce, N_firm - V_require, N_firm - V_take, N_firm - V_use, N_flow - N_information, N_food - J_available, N_food - V_get, N_food - V_produce, N_food - V_share, N_force - N_use, N_form - J_common, N_form - J_new, N_form - V_take, N_form - V_use, N_framework - V_develop, N_generation - J_new, N_goal - J_common, N_good - V_produce, N_government - V_make, N_government - V_need, N_government - V_require, N_government - V_take, N_group - V_make, N_group - V_use, N_hand - V_use, N_help - V_need, N_hypothesis - V_develop, N_idea - J_new, N_idea - V_apply, N_idea - V_develop, N_idea - V_get, N_idea - V_give, N_idea - V_share, N_immigrant - J_new, N_incentive - V_give, N_income - V_give, N_individual - V_allow, N_individual - V_give, N_individual - V_learn, N_individual - V_make, N_individual - V_share, N_industry - J_new, N_inference - V_make, N_information - J_available, N_information - J_new, N_information - N_information, N_information - V_get, N_information - V_give, N_information - V_learn, N_information - V_make, N_information - V_need, N_information - V_share, N_information - V_use, N_innovation - J_new, N_introduction - J_new, N_investment - J_new, N_investment - V_make, N_investment - V_require, N_issue - J_new, N_job - J_available, N_job - J_new, N_job - V_get, N_job - V_require, N_job - V_take, N_judgment - V_make, N_kind - J_new, N_kind - N_information, N_knowledge - J_new, N_knowledge - V_apply, N_knowledge - V_require, N_knowledge - V_share, N_knowledge - V_use, N_labor - V_produce, N_labor - V_require, N_land - N_use, N_language - J_common, N_language - N_use, N_language - V_develop, N_language - V_learn, N_language - V_use, N_law - J_new, N_law - V_allow, N_law - V_apply, N_law - V_require, N_level - V_give, N_level - V_require, N_life - V_make, N_line - J_new, N_living - V_make, N_loan - V_make, N_location - J_new, N_logic - V_apply, N_look - V_take, N_lot - V_learn, N_market - J_new, N_market - N_information, N_market - V_develop, N_market - V_make, N_material - J_available, N_material - N_use, N_material - V_learn, N_material - V_use, N_measure - J_available, N_measure - N_use, N_measure - V_develop, N_measure - V_take, N_measure - V_use, N_member - J_new, N_member - V_share, N_memory - N_information, N_method - J_new, N_method - N_use, N_method - V_allow, N_method - V_apply, N_method - V_develop, N_method - V_use, N_model - V_allow, N_model - V_apply, N_model - V_develop, N_model - V_use, N_money - V_get, N_money - V_give, N_money - V_make, N_money - V_need, N_money - V_take, N_money - V_use, N_month - V_take, N_name - V_give, N_nation - V_develop, N_note - V_take, N_number - V_give, N_number - V_require, N_number - V_use,

N_observation - V_learn, N_observation - V_make, N_observation - V_use, N_offspring - V_produce, N_one - J_new, N_opportunity - J_available, N_opportunity - J_new, N_opportunity - V_give, N_opportunity - V_learn, N_option - J_available, N_order - V_get, N_organization - J_new, N_other - V_learn, N_other - V_make, N_other - V_share, N_outcome - V_produce, N_output - V_produce, N_part - V_take, N_participant - V_give, N_participant - V_learn, N_participant - V_require, N_pattern - J_common, N_pattern - N_use, N_payment - V_make, N_people - V_allow, N_people - V_get, N_people - V_give, N_people - V_learn, N_people - V_make, N_people - V_share, N_people - V_take, N_people - V_use, N_period - V_give, N_person - V_get, N_perspective - V_take, N_piece - N_information, N_place - V_take, N_plan - V_develop, N_plant - J_new, N_plant - V_produce, N_point - V_give, N_point - V_make, N_policy - V_make, N_position - V_take, N_possibility - V_allow, N_power - V_give, N_practice - J_common, N_practice - J_new, N_practice - V_use, N_prediction - V_make, N_price - V_give, N_price - V_use, N_principle - V_apply, N_problem - J_new, N_procedure - V_use, N_process - N_information, N_process - V_learn, N_process - V_make, N_process - V_produce, N_process - V_use, N_processing - N_information, N_processing - V_require, N_product - J_new, N_product - V_develop, N_product - V_make, N_product - V_produce, N_profit - V_make, N_program - V_develop, N_progress - V_make, N_project - V_require, N_project - V_take, N_purchase - V_make, N_question - V_need, N_rate - V_give, N_rate - V_require, N_rate - V_use, N_regression - V_use, N_relationship - V_develop, N_research - J_new, N_research - V_develop, N_research - V_need, N_research - V_use, N_researcher - V_allow, N_researcher - V_develop, N_researcher - V_need, N_researcher - V_use, N_resource - J_available, N_resource - J_new, N_resource - N_use, N_resource - V_need, N_resource - V_require, N_resource - V_use, N_responsibility - V_take, N_result - V_produce, N_result - V_use, N_right - V_give, N_rise - V_give, N_risk - V_take, N_role - V_take, N_rule - V_apply, N_rule - V_learn, N_sample - V_give, N_sample - V_take, N_sample - V_use, N_science - V_apply, N_sense - J_common, N_sense - V_develop, N_sense - V_make, N_service - J_available, N_service - V_produce, N_set - J_common, N_set - J_new, N_set - V_give, N_share - J_common, N_situation - V_give, N_skill - V_develop, N_skill - V_learn, N_skill - V_need, N_skill - V_require, N_something - V_get, N_something - V_learn, N_source - J_available, N_source - J_new, N_source - N_information, N_source - N_use, N_specie - J_new, N_statement - V_make, N_statistic - V_use, N_step - V_take, N_strategy - J_new, N_strategy - N_use, N_strategy - V_develop, N_strategy - V_use, N_structure - N_information, N_student - V_give, N_student - V_learn, N_student - V_require, N_student - V_take, N_study - V_use, N_system - N_information, N_system - V_allow, N_system - V_develop, N_system - V_make, N_system - V_produce, N_system - V_use, N_task - V_give, N_task - V_learn, N_task - V_require, N_technique - J_new, N_technique - N_use, N_technique - V_apply, N_technique - V_develop, N_technique - V_learn, N_technique - V_use, N_technology - J_available, N_technology - J_new, N_technology - N_information, N_technology - N_use, N_technology - V_allow, N_technology - V_develop, N_technology - V_use, N_term - N_use, N_term - V_apply, N_term - V_use, N_test - V_take, N_test - V_use, N_theme - J_common, N_theory - J_new, N_theory - V_apply, N_theory - V_develop, N_theory - V_learn, N_thing - V_get, N_thing - V_learn, N_thing - V_make, N_thing - V_need, N_time - V_get, N_time - V_give, N_time - V_make, N_time - V_require, N_time - V_take, N_time - V_use, N_tool - J_new, N_tool - N_use, N_tool - V_develop, N_tool - V_make, N_tool - V_use, N_training - V_require, N_treatment - V_require, N_type - N_information, N_type - V_use, N_understanding - V_require, N_unit - V_produce, N_use - J_common, N_use - V_make, N_use - V_require, N_value - V_give, N_value - V_take, N_value - V_use, N_variable - V_use, N_variety - V_use, N_view - V_share, N_view - V_take, N_way - J_new, N_way - V_develop, N_way - V_get, N_way - V_give, N_way - V_learn, N_way - V_make, N_way - V_use, N_weight - V_give, N_woman - V_give, N_woman - V_make, N_word - V_learn, N_word - V_use, N_work - V_need, N_worker - V_produce, N_world - V_develop, N_year - V_give, N_year - V_take, V_acquire - J_new, V_acquire - N_information, V_add - J_new, V_address - V_need, V_adopt - J_new, V_allow - J_new, V_allow - V_develop, V_base - N_information, V_become - J_available, V_become - J_common, V_begin - J_new, V_bring - J_new, V_build - J_new, V_buy - V_need, V_calculate - V_use, V_call - J_new, V_collect - N_information, V_come - J_new, V_communicate - N_information, V_compare - V_allow, V_consider - V_need, V_contain - N_information, V_create - J_new, V_date - V_use, V_decide - V_need, V_describe - V_use, V_determine - V_need, V_determine - V_use, V_develop - J_new, V_develop - V_allow, V_distinguish - V_allow, V_distribute - V_produce, V_emerge - J_new, V_encode - N_information, V_enter - J_new, V_estimate - V_use, V_expect - V_give, V_extract - N_information, V_feel - V_make, V_find - J_new, V_find -

V_use, V_form - J_new, V_gather - N_information, V_generate - J_new, V_get - N_information, V_get - V_need, V_give - J_new, V_give - N_information, V_give - V_take, V_go - V_get, V_grant - V_take, V_help - N_information, V_help - V_develop, V_help - V_make, V_hold - N_information, V_identify - V_allow, V_identify - V_use, V_include - J_new, V_include - N_information, V_increase - N_use, V_integrate - N_information, V_introduce - J_new, V_involve - N_information, V_involve - N_use, V_keep - V_need, V_know - V_get, V_know - V_need, V_lead - J_new, V_learn - J_new, V_learn - N_information, V_learn - V_need, V_learn - V_share, V_learn - V_use, V_limit - N_use, V_look - V_need, V_make - J_available, V_make - J_new, V_make - N_information, V_make - N_use, V_make - V_need, V_make - V_use, V_measure - V_use, V_need - J_new, V_need - N_information, V_need - V_get, V_need - V_learn, V_need - V_make, V_obtain - N_information, V_obtain - V_use, V_offer - J_new, V_open - J_new, V_pay - V_get, V_pay - V_need, V_pay - V_require, V_present - N_information, V_process - N_information, V_produce - J_new, V_produce - V_use, V_promote - N_use, V_provide - J_new, V_provide - N_information, V_raise - J_new, V_receive - N_information, V_regard - N_information, V_require - J_new, V_require - N_use, V_reveal - N_information, V_seem - V_make, V_sell - J_new, V_sell - V_produce, V_share - J_common, V_share - N_information, V_share - V_learn, V_solve - V_need, V_speak - V_learn, V_start - J_new, V_start - V_get, V_study - V_learn, V_study - V_use, V_support - V_need, V_take - J_new, V_take - V_give, V_test - V_allow, V_test - V_develop, V_test - V_use, V_try - V_get, V_try - V_make, V_understand - V_need, V_use - N_information, V_use - V_learn, V_use - V_make, V_use - V_produce, V_want - V_get, V_work - V_need
Collocational Dimension 10: Social Welfare (638 collocations)
J_additional - N_support, J_additional - V_provide, J_available - N_job, J_available - N_material, J_available - N_opportunity, J_available - N_resource, J_available - N_service, J_bad - J_good, J_central - N_government, J_current - N_policy, J_detailed - V_provide, J_direct - N_access, J_direct - N_benefit, J_direct - N_support, J_direct - V_provide, J_domestic - N_service, J_easy - N_access, J_economic - N_benefit, J_economic - N_government, J_economic - N_opportunity, J_economic - N_policy, J_economic - N_resource, J_educational - N_opportunity, J_educational - N_program, J_empirical - N_support, J_empirical - V_provide, J_federal - N_government, J_federal - N_policy, J_few - N_job, J_few - N_opportunity, J_few - N_resource, J_final - N_good, J_financial - N_policy, J_financial - N_resource, J_financial - N_service, J_financial - N_support, J_financial - V_provide, J_foreign - J_direct, J_foreign - N_government, J_free - N_access, J_free - V_provide, J_further - N_support, J_further - V_provide, J_global - N_health, J_good - N_access, J_good - N_health, J_good - N_job, J_good - N_opportunity, J_good - N_service, J_good - V_provide, J_great - N_benefit, J_great - N_care, J_great - N_opportunity, J_great - N_resource, J_great - V_provide, J_high - N_job, J_human - N_health, J_human - N_resource, J_important - N_policy, J_important - V_provide, J_indirect - J_direct, J_international - J_financial, J_limited - N_access, J_limited - N_opportunity, J_limited - N_resource, J_local - N_government, J_long-term - N_benefit, J_low - N_benefit, J_low - N_job, J_many - N_government, J_medical - N_care, J_medical - N_health, J_medical - N_service, J_mental - N_health, J_military - N_service, J_monetary - N_policy, J_more - J_good, J_more - N_resource, J_more - V_provide, J_national - N_government, J_national - N_policy, J_natural - N_resource, J_necessary - N_resource, J_necessary - V_provide, J_new - N_job, J_new - N_opportunity, J_new - N_resource, J_new - V_provide, J_other - N_good, J_other - N_resource, J_own - N_government, J_particular - N_policy, J_physical - N_health, J_political - N_resource, J_political - N_support, J_poor - N_health, J_potential - N_benefit, J_primary - N_care, J_private - J_public, J_private - N_government, J_public - N_good, J_public - N_government, J_public - N_health, J_public - N_policy, J_public - N_program, J_public - N_resource, J_public - N_service, J_public - N_support, J_public - V_provide, J_raw - N_material, J_real - J_financial, J_rich - N_resource, J_scarce - N_resource, J_social - N_policy, J_social - N_resource, J_social - N_service, J_social - N_support, J_social - V_provide, J_strong - N_support, J_strong - V_provide, J_such - J_public, J_such - N_material, J_such - N_policy, J_such - N_program, J_such - N_resource, J_such - N_service, J_such - V_provide, J_sufficient - N_resource, J_useful - V_provide, J_valuable - N_resource, J_various - N_government, J_various - N_policy, N_access - J_direct, N_access - J_good, N_access - N_care, N_access - N_health, N_access - N_resource, N_access - N_service, N_access - V_provide, N_action - N_government, N_advantage - N_opportunity, N_agency - N_government, N_allocation - N_resource, N_amount - N_resource, N_analysis - N_policy, N_analysis - V_provide, N_answer - V_provide, N_artifact - N_material, N_asset - J_financial,

N_availability - N_resource, N_awareness - J_public, N_bank - J_financial, N_bank - N_government, N_basis - V_provide, N_benefit - J_direct, N_benefit - N_health, N_benefit - V_provide, N_business - N_government, N_business - N_resource, N_capital - J_financial, N_care - N_access, N_care - N_health, N_care - N_service, N_care - V_provide, N_challenge - N_opportunity, N_chance - J_good, N_change - N_policy, N_child - N_care, N_child - N_health, N_choice - J_good, N_community - N_health, N_company - J_financial, N_company - J_good, N_company - V_provide, N_comparison - J_direct, N_competition - N_resource, N_computer - N_program, N_concern - J_public, N_concern - N_health, N_condition - J_good, N_condition - N_material, N_connection - J_direct, N_consequence - N_health, N_consumer - N_good, N_consumer - N_service, N_consumption - N_good, N_contact - J_direct, N_control - J_direct, N_control - N_government, N_control - N_resource, N_cost - J_financial, N_cost - N_benefit, N_cost - N_care, N_cost - N_good, N_cost - N_health, N_cost - N_opportunity, N_crisis - J_financial, N_culture - N_material, N_customer - N_service, N_datum - V_provide, N_deal - J_good, N_debate - J_public, N_decision - J_financial, N_decision - J_good, N_decision - N_policy, N_delivery - N_service, N_description - V_provide, N_development - N_program, N_disease - N_health, N_display - J_public, N_distribution - N_good, N_distribution - N_resource, N_economy - N_government, N_economy - N_policy, N_education - J_public, N_education - N_access, N_education - N_care, N_education - N_health, N_education - N_job, N_education - N_program, N_effect - J_direct, N_effect - N_policy, N_effort - J_good, N_employee - N_job, N_employment - N_job, N_employment - N_opportunity, N_energy - N_resource, N_environment - V_provide, N_estimate - J_good, N_estimate - V_provide, N_evidence - J_direct, N_evidence - J_good, N_evidence - N_support, N_evidence - V_provide, N_example - J_good, N_example - V_provide, N_exchange - N_good, N_exchange - N_service, N_explanation - V_provide, N_factory - N_job, N_family - N_health, N_family - N_support, N_feedback - V_provide, N_finding - V_provide, N_firm - J_financial, N_firm - J_good, N_firm - N_resource, N_fit - J_good, N_food - N_access, N_food - N_good, N_food - N_resource, N_food - V_provide, N_framework - V_provide, N_friend - J_good, N_fund - V_provide, N_goal - N_policy, N_good - J_public, N_good - N_material, N_good - N_service, N_good - V_provide, N_government - J_public, N_government - N_policy, N_government - N_program, N_government - N_service, N_government - N_support, N_government - V_provide, N_grade - J_good, N_health - J_good, N_health - J_public, N_health - N_access, N_health - N_benefit, N_health - N_care, N_health - N_program, N_health - N_service, N_health - V_provide, N_hour - N_job, N_housing - J_public, N_hypothesis - N_support, N_idea - J_good, N_idea - N_support, N_impact - J_direct, N_impact - N_health, N_impact - N_policy, N_implication - N_policy, N_incentive - V_provide, N_indicator - N_health, N_industry - N_service, N_information - V_provide, N_insight - V_provide, N_institution - J_financial, N_institution - J_public, N_institution - N_government, N_insurance - N_benefit, N_insurance - N_health, N_insurance - N_policy, N_insurance - V_provide, N_interest - J_good, N_interest - J_public, N_interest - N_government, N_interest - N_policy, N_internet - N_access, N_intervention - N_government, N_investment - J_direct, N_investment - J_financial, N_investment - J_good, N_investment - N_opportunity, N_issue - J_public, N_issue - N_health, N_issue - N_policy, N_job - J_good, N_job - N_opportunity, N_job - N_service, N_job - V_provide, N_labor - N_good, N_labor - N_material, N_labor - N_resource, N_lack - N_support, N_land - N_resource, N_law - N_government, N_law - N_policy, N_life - J_good, N_life - N_health, N_link - J_direct, N_loss - J_financial, N_loss - N_job, N_maker - N_policy, N_management - J_good, N_management - N_resource, N_manufacturing - N_job, N_market - J_financial, N_market - N_good, N_market - N_opportunity, N_material - N_good, N_material - N_resource, N_material - V_provide, N_means - V_provide, N_measure - J_direct, N_measure - J_good, N_measure - V_provide, N_memory - J_good, N_model - J_good, N_model - V_provide, N_money - N_good, N_money - N_government, N_news - J_good, N_object - N_material, N_observation - J_direct, N_offer - N_job, N_official - N_government, N_opinion - J_public, N_opportunity - J_good, N_opportunity - N_job, N_opportunity - V_provide, N_outcome - J_good, N_outcome - N_health, N_people - J_good, N_people - V_provide, N_performance - J_good, N_performance - N_job, N_place - J_good, N_place - J_public, N_plan - N_health, N_policy - J_financial, N_policy - J_public, N_policy - N_government, N_policy - N_program, N_population - N_health, N_power - N_government, N_power - N_resource, N_practice - J_good, N_practice - N_policy, N_price - J_good, N_price - N_good, N_problem - N_health, N_product - J_good, N_product - N_service, N_production - N_good, N_professional - N_health, N_program - J_public, N_program - N_government, N_program - N_health, N_program - N_policy, N_program - N_support, N_program - V_provide, N_project - J_public, N_protection - V_provide,

N_provider - N_care, N_provider - N_health, N_provider - N_service, N_quality - J_good, N_quality - N_care, N_rate - N_policy, N_reason - J_good, N_record - V_provide, N_regulation - N_government, N_relation - J_public, N_resource - J_financial, N_resource - J_public, N_resource - N_access, N_resource - N_material, N_resource - N_support, N_resource - V_provide, N_response - J_good, N_response - N_policy, N_result - J_good, N_result - V_provide, N_risk - J_financial, N_risk - N_health, N_role - N_government, N_rule - N_policy, N_safety - N_health, N_sale - N_good, N_school - J_public, N_school - N_program, N_search - N_job, N_section - V_provide, N_sector - J_financial, N_sector - J_public, N_sector - N_job, N_sector - N_service, N_security - N_job, N_security - V_provide, N_service - J_financial, N_service - J_good, N_service - J_public, N_service - N_access, N_service - N_care, N_service - N_good, N_service - N_government, N_service - N_health, N_service - N_job, N_service - V_provide, N_skill - N_job, N_source - N_material, N_source - V_provide, N_space - J_public, N_spending - N_government, N_state - N_government, N_state - N_policy, N_status - N_health, N_strategy - J_good, N_study - V_provide, N_subsidy - N_government, N_support - J_direct, N_support - J_financial, N_support - J_public, N_support - N_government, N_support - N_program, N_support - N_resource, N_support - V_provide, N_system - J_financial, N_system - J_public, N_system - N_care, N_system - N_health, N_table - V_provide, N_tax - N_benefit, N_tax - N_government, N_tax - N_policy, N_technology - N_access, N_television - N_program, N_test - V_provide, N_thing - J_good, N_time - J_good, N_tool - N_material, N_trade - N_good, N_trade - N_policy, N_training - N_job, N_training - N_program, N_training - V_provide, N_treatment - N_care, N_treatment - N_health, N_type - N_job, N_understanding - J_good, N_use - N_material, N_use - N_resource, N_value - N_good, N_wage - N_benefit, N_wage - N_job, N_water - N_resource, N_way - J_good, N_way - V_provide, N_welfare - N_policy, N_woman - N_health, N_work - N_job, N_worker - N_benefit, N_worker - N_job, N_worker - N_service, V_achieve - J_good, V_acquire - N_resource, V_adopt - N_policy, V_affect - N_health, V_affect - N_policy, V_aim - N_policy, V_aim - N_program, V_allocate - N_resource, V_argue - N_policy, V_associate - N_benefit, V_associate - N_health, V_believe - J_good, V_bring - N_benefit, V_buy - N_good, V_change - N_policy, V_choose - J_good, V_consider - N_policy, V_control - N_resource, V_create - N_government, V_create - N_job, V_create - N_opportunity, V_create - N_policy, V_create - N_program, V_design - N_policy, V_design - N_program, V_develop - N_program, V_discuss - N_policy, V_ensure - N_access, V_exist - N_opportunity, V_exploit - N_opportunity, V_feel - J_good, V_find - J_good, V_find - N_job, V_find - N_support, V_follow - N_policy, V_gain - N_access, V_gain - N_benefit, V_gain - N_support, V_get - J_good, V_get - N_job, V_give - J_good, V_give - N_opportunity, V_help - N_policy, V_help - N_program, V_help - V_provide, V_hold - N_job, V_implement - N_policy, V_implement - N_program, V_impose - N_government, V_improve - J_financial, V_improve - N_care, V_improve - N_health, V_include - N_health, V_include - N_material, V_include - N_resource, V_increase - J_financial, V_increase - N_benefit, V_increase - N_government, V_influence - J_public, V_influence - N_policy, V_keep - N_job, V_know - J_good, V_lead - N_policy, V_learn - N_material, V_learn - N_opportunity, V_leave - N_job, V_lend - N_support, V_limit - N_access, V_limit - N_opportunity, V_limit - N_resource, V_look - J_good, V_look - N_job, V_lose - N_job, V_make - J_good, V_make - N_government, V_make - N_policy, V_maximize - N_benefit, V_miss - N_opportunity, V_move - N_resource, V_need - N_government, V_need - N_resource, V_offer - J_good, V_offer - N_benefit, V_offer - N_job, V_offer - N_opportunity, V_offer - N_support, V_pay - N_benefit, V_pay - N_government, V_pay - N_job, V_perform - J_good, V_perform - N_job, V_present - N_opportunity, V_produce - J_good, V_produce - N_good, V_produce - N_service, V_promote - N_health, V_promote - N_policy, V_protect - N_resource, V_provide - J_direct, V_provide - J_financial, V_provide - J_good, V_provide - J_public, V_provide - N_access, V_provide - N_benefit, V_provide - N_care, V_provide - N_good, V_provide - N_government, V_provide - N_health, V_provide - N_job, V_provide - N_material, V_provide - N_opportunity, V_provide - N_program, V_provide - N_resource, V_provide - N_service, V_provide - N_support, V_purchase - N_good, V_pursue - N_policy, V_realize - N_benefit, V_receive - N_benefit, V_receive - N_care, V_receive - N_health, V_receive - N_job, V_receive - N_service, V_receive - N_support, V_reduce - N_benefit, V_reduce - N_government, V_reduce - N_policy, V_reduce - N_program, V_relate - N_health, V_require - J_good, V_require - N_government, V_require - N_job, V_require - N_resource, V_run - N_government, V_run - N_program, V_seek - N_opportunity, V_sell - N_good, V_sell - N_service, V_set - N_government, V_set - N_policy, V_support - N_resource, V_take - N_care, V_take - N_government, V_take - N_job, V_think - J_good, V_trade - N_good, V_treat - N_health, V_use - N_material, V_use - N_resource, V_work - J_good, V_work - N_job

Collocational Dimension 11: Marking Time and Sectioning (410 collocations)
J_alternative - V_consider, J_average - N_day, J_critical - N_period, J_current - N_period, J_different - J_several, J_different - N_period, J_early - N_century, J_early - N_chapter, J_early - N_day, J_early - N_period, J_early - N_year, J_early - V_begin, J_entire - N_period, J_entire - V_spend, J_european - N_century, J_few - J_first, J_few - J_next, J_few - N_century, J_few - N_day, J_few - N_year, J_final - N_section, J_first - J_few, J_first - N_century, J_first - N_day, J_first - N_period, J_first - N_year, J_first - V_consider, J_following - N_section, J_important - J_several, J_important - V_consider, J_key - J_several, J_last - J_few, J_last - J_several, J_last - N_century, J_last - N_chapter, J_last - N_day, J_last - N_period, J_last - N_section, J_last - N_year, J_late - N_century, J_late - N_chapter, J_late - N_year, J_late - V_begin, J_least - N_year, J_less - N_year, J_less - V_spend, J_little - V_spend, J_long - N_period, J_low - J_long, J_main - N_section, J_major - J_several, J_many - N_year, J_more - J_few, J_more - N_century, J_more - N_year, J_more - V_discuss, J_more - V_spend, J_most - V_spend, J_much - V_spend, J_new - V_begin, J_next - J_few, J_next - J_several, J_next - N_century, J_next - N_chapter, J_next - N_day, J_next - N_period, J_next - N_section, J_next - N_year, J_next - V_consider, J_next - V_discuss, J_nineteenth - N_century, J_old - N_year, J_other - J_several, J_past - J_few, J_past - J_several, J_past - N_century, J_past - N_year, J_previous - N_chapter, J_previous - N_period, J_previous - N_section, J_previous - N_year, J_previous - V_discuss, J_recent - N_year, J_same - N_period, J_same - N_year, J_second - J_first, J_second - N_section, J_second - N_year, J_several - J_next, J_several - N_day, J_several - N_year, J_short - J_long, J_short - N_period, J_simple - V_consider, J_theoretical - N_section, J_twentieth - J_first, J_twentieth - N_century, J_various - V_discuss, N_advantage - J_several, N_age - N_year, N_amount - V_spend, N_analysis - V_begin, N_aspect - J_several, N_average - N_day, N_beginning - N_century, N_beginning - N_chapter, N_beginning - N_period, N_billion - N_year, N_billion - V_spend, N_book - N_chapter, N_book - V_discuss, N_career - V_begin, N_case - V_consider, N_century - J_few, N_century - J_first, N_century - J_next, N_century - V_begin, N_chapter - J_next, N_chapter - V_begin, N_chapter - V_consider, N_chapter - V_discuss, N_characteristic - J_several, N_column - J_first, N_company - V_begin, N_concept - V_discuss, N_condition - J_first, N_country - J_few, N_country - J_several, N_datum - N_year, N_day - J_few, N_day - J_first, N_day - J_next, N_day - J_several, N_day - N_period, N_day - V_spend, N_decade - J_few, N_decade - J_first, N_decade - J_next, N_decade - J_several, N_decade - N_century, N_decision - V_consider, N_detail - J_next, N_detail - N_chapter, N_detail - N_section, N_detail - V_discuss, N_difference - J_first, N_discussion - N_chapter, N_discussion - N_section, N_distance - J_long, N_dollar - J_few, N_dollar - V_spend, N_duration - J_long, N_education - N_year, N_effect - V_consider, N_end - N_century, N_end - N_chapter, N_end - N_day, N_end - N_period, N_end - N_year, N_evidence - J_first, N_example - V_consider, N_exception - J_few, N_experiment - J_first, N_factor - J_several, N_factor - V_consider, N_feature - J_several, N_finding - V_discuss, N_firm - V_consider, N_following - N_section, N_following - V_consider, N_generation - J_next, N_group - J_several, N_growth - N_period, N_half - J_first, N_half - N_century, N_history - J_long, N_history - N_period, N_hour - J_few, N_hour - J_long, N_hour - N_day, N_hour - V_spend, N_implication - V_discuss, N_income - V_spend, N_individual - J_few, N_interest - N_year, N_interval - J_long, N_issue - V_consider, N_issue - V_discuss, N_job - J_few, N_kind - J_several, N_lag - J_long, N_length - N_period, N_life - J_long, N_life - N_year, N_line - J_long, N_list - J_long, N_lot - V_spend, N_marriage - N_year, N_method - J_several, N_method - V_discuss, N_middle - N_century, N_million - N_year, N_minute - J_few, N_model - N_section, N_model - V_consider, N_money - V_spend, N_month - J_few, N_month - J_first, N_month - J_next, N_month - J_several, N_month - N_day, N_month - N_year, N_month - V_spend, N_name - J_first, N_night - N_day, N_night - V_spend, N_number - N_day, N_number - N_period, N_number - N_year, N_observation - J_first, N_opportunity - J_few, N_order - J_first, N_page - J_next, N_part - J_first, N_people - J_few, N_people - V_consider, N_percent - N_year, N_period - J_first, N_period - J_long, N_period - J_next, N_period - N_day, N_period - N_year, N_period - V_begin, N_person - J_first, N_phase - J_first, N_place - J_first, N_point - V_discuss, N_policy - V_consider, N_policy - V_discuss, N_possibility - V_consider, N_problem - V_consider, N_problem - V_discuss, N_process - V_begin, N_quarter - J_first, N_question - J_first, N_question - V_consider, N_rate - N_period, N_rate - N_year, N_reason - J_several, N_reason - V_discuss, N_recall - N_chapter, N_research - V_begin, N_researcher - V_begin, N_resource - J_few, N_result - N_section, N_run - J_long, N_sample - N_period, N_school - N_day, N_school - N_year, N_section - J_next, N_section - N_chapter, N_section - V_consider, N_section - V_discuss, N_series - J_long,

N_stage - J_first, N_stage - J_next, N_step - J_first, N_step - J_next, N_study - J_few, N_study - J_first, N_study - J_several, N_study - V_consider, N_technique - V_discuss, N_term - J_first, N_term - J_long, N_text - V_discuss, N_thing - J_first, N_thousand - N_year, N_time - J_first, N_time - J_long, N_time - J_several, N_time - N_day, N_time - N_period, N_time - N_year, N_time - V_spend, N_topic - N_chapter, N_topic - V_discuss, N_tradition - J_long, N_turn - N_century, N_way - J_long, N_way - J_several, N_week - J_few, N_week - J_first, N_week - J_next, N_week - J_several, N_week - N_day, N_week - V_spend, N_work - V_consider, N_worker - J_few, N_year - J_few, N_year - J_first, N_year - J_next, N_year - J_several, N_year - N_period, N_year - V_begin, N_year - V_spend, V_analyze - N_section, V_appear - J_first, V_become - J_first, V_begin - N_century, V_begin - N_chapter, V_begin - N_period, V_begin - N_section, V_begin - N_year, V_begin - V_consider, V_come - J_first, V_come - N_year, V_compare - N_day, V_conclude - N_section, V_consider - J_first, V_consider - J_next, V_consider - N_chapter, V_consider - N_section, V_consider - V_begin, V_contain - J_few, V_contain - J_several, V_cover - N_chapter, V_date - N_year, V_decline - V_begin, V_define - N_chapter, V_describe - J_next, V_describe - N_chapter, V_describe - N_section, V_develop - N_chapter, V_discuss - J_next, V_discuss - N_chapter, V_discuss - N_section, V_earn - N_year, V_emerge - V_begin, V_end - V_begin, V_examine - J_few, V_examine - J_next, V_examine - N_chapter, V_examine - N_section, V_explain - N_chapter, V_explore - N_chapter, V_explore - N_section, V_find - J_several, V_fix - N_period, V_focus - N_chapter, V_focus - N_section, V_follow - J_first, V_follow - N_period, V_follow - N_section, V_give - N_day, V_give - N_period, V_give - N_year, V_go - J_long, V_grow - N_year, V_happen - V_consider, V_illustrate - N_section, V_illustrate - V_consider, V_include - J_several, V_introduce - N_chapter, V_introduce - N_section, V_learn - N_chapter, V_leave - J_few, V_let - V_begin, V_let - V_consider, V_live - J_long, V_live - N_year, V_look - N_chapter, V_look - N_section, V_mention - N_chapter, V_move - J_next, V_name - J_few, V_need - V_consider, V_note - N_chapter, V_occur - J_first, V_occur - N_period, V_pay - N_year, V_present - N_chapter, V_present - N_section, V_produce - J_few, V_provide - N_section, V_read - N_chapter, V_recall - N_chapter, V_remain - J_few, V_review - N_section, V_see - J_next, V_see - N_chapter, V_see - N_section, V_seem - J_first, V_sell - V_begin, V_show - J_several, V_show - N_chapter, V_show - N_section, V_spend - N_day, V_spend - N_year, V_study - N_chapter, V_study - V_spend, V_summarize - N_section, V_take - J_long, V_take - N_day, V_take - N_year, V_turn - J_next, V_understand - V_begin, V_work - J_few, V_work - J_long, V_work - N_day

Nodes and collocates are preceded by their part-of-speech. J=Adjective; N=Noun; V=Verb