Multi-dimensional Analysis, Text Constellations, and Interdisciplinary Discourse

Abstract
Although corpus research, and multi-dimensional analysis, has long been used to investigate the discourse of academic research, it is usual to compare strictly discrete disciplines and to focus on that which is maximally distinctive about each discipline. The project reported in this paper instead investigates an interdisciplinary journal (Global Environmental Change) and establishes Text Constellations comprising papers from the journal that share a dimensional profile. The dimensions are established by applying the multidimensional method to the Birmingham-Elsevier Environment Corpus, a 50 million word corpus consisting of papers from 11 journals. The corpus was tagged using the Biber tagger, and factor analysis was used to identify six dimensions that are unique to this corpus. The paper discusses the statistical and phraseological characteristics of six Text Constellations. It argues that the concepts of divergence and convergence are useful in conceptualizing interdisciplinary research discourse.

1. Introduction

The existence of so-called academic 'silos' (e.g. Stirling 2014) has often been seen as a barrier to research that will effectively solve pressing global problems, and the growth of interdisciplinary fields that combine types of expertise in addressing such problems is welcomed. To a large extent, however, applied corpus linguistic research into academic discourse has lagged behind this development by focusing primarily on identifiable and discrete disciplines (e.g. Hyland 2000; Charles 2006; Fløttum (ed.) 2007; Hyland & Bondi (eds.) 2006). The project reported in this paper aims to challenge this focus by investigating the language of an interdisciplinary field.[1] Instead of taking as a starting point sub-corpora composed of texts from contrasting disciplines, our corpus consists of papers selected irrespective of discipline but published in journals dealing with particular topics. The aim is to find ways of describing the discourse of interdisciplinary research. As there can be no more pressing contemporary problem than that of human-made environmental change, we selected the journal Global Environmental Change as the main target of the study. This is a self-declared interdisciplinary journal, identified as successful by its publisher, that focuses on the 'human and policy dimensions of global environmental change'.[2] Alongside this journal, papers from a further 10 journals, all dealing with environmental studies and related topics, are included in the corpus, shown in Section 2. We selected multidimensional analysis as a key component of our research, but applied it in a novel way, as explained below.

The technique of multidimensional analysis (Biber, Conrad, Reppen, Byrd & Helt 1988, 1995) as a means of investigating register variation has an extensive history in corpus linguistics. It has been employed in various areas of research including ESP studies (Biber, Conrad, & Reppen, 1998), first and second language acquisition (Biber et al., 1998), diachronic studies of stylistics (Biber et al., 1998), language testing (Connor-Linton & Shohamy, 2001), discourse analysis on university registers (Biber, Csomay, Jones, & Keck, 2004; Csomay, 2007; Gardner, Biber and Nesi 2015), taxonomies of web pages (Biber &

[1] The project was funded by the ESRC (project number ES/K007300/1). The title of the project is Interdisciplinary Research Discourse: The case of Global Environmental Change.
[2] http://www.journals.elsevier.com/global-environmental-change/

Kurjian, 2007), and rhetorical moves in research articles (Kanoksilapatham, 2007). At least five of the seven dimensions of register variation originally proposed by Biber (1988) are often still called upon to account for diachronic, regional, or functional difference.

In common with Biber et al (2002), Biber et al. (2004), Biber (2006) and Gray (2013), among other studies, the project reported in this paper uses multidimensional analysis (MDA) to investigate academic discourse in English, but it involves a number of innovations. The corpus used is relatively specific, consisting of only written texts, only research papers published in international academic journals, and only papers about environmental change and associated fields. It has therefore been necessary to derive factors and dimensions that are specific to the corpus, rather than replicating the dimensions proposed by Biber and others based on a wider range of registers. Following the identification of six dimensions, two methods of corpus comparison are then used. One is traditional: a number of sub-corpora identified on external criteria are compared in terms of the dimensions, so that the extent of difference or similarity can be observed. In this case, each sub-corpus comprises the papers in a given academic journal, the criterion for inclusion of a text in the sub-corpus being that it has been published in that journal. This comparison allows us to identify what might be called the 'house style' of each journal. The second method is more novel: investigating only one journal, Global Environmental Change, clusters of papers that demonstrate similarity in terms of their weighting on each dimension are identified (Biber 1989). Each of these clusters, which we have called 'constellations' of papers or 'Text Constellations',[3] comprises a sub-corpus identified on internal criteria, that is, by the similarity of their dimension profiles. The 'constellations' are then scrutinised to identify the discourse features that are indexed by the relevant dimension profiles. Some preliminary phraseological information is also obtained, using the SketchEngine software.[4] We hope to use the distinctive discourse features of the journal to inform our understanding of interdisciplinary research practices and identify how researchers tackle the challenge of writing for a cross-disciplinary audience.

More specifically, the research questions to be answered in this paper are as follows:

1. Using a corpus of a high level of specificity (written articles from a narrow range of academic journals), what factors are identified by multidimensional analysis and how might they be interpreted as dimensions?
2. How similar are the papers in each of the 11 journals (Global Environmental Change and 10 others) in our corpus? And how different are the journals from each other? In other words, to what extent do the dimensions distinguish between individual journals? To the extent that individual journals are distinguished by dimensions, is it also the case that interdisciplinary journals can be distinguished from monodisciplinary ones?

[3] We are not alone in appropriating the 'constellation' metaphor. See, for example, Cantos & Sánchez (2001).

[4] In common with many others, we are grateful to the generosity of spirit of Adam Kilgariff, who volunteered SketchEngine for our use.

3. Focusing on Global Environmental Change only, when individual texts are aligned with each other in terms of dimension scores, what constellations of papers can be identified and how can these be described? What can phraseological studies add to these descriptions?

At the end of the paper we discuss also what the implications of the identification of Text Constellations might be for the practice of writing in interdisciplinary fields.

In the sections below, we describe the corpus we are using (section 2), outline the process of identifying the six dimensions (section 3), present and describe the six Text Constellations in the journal Global Environmental Change (section 4). Finally, we present an interpretation of our results in terms of the nature of interdisciplinary discourse and the practice of corpus linguistics (section 5).

2. The Birmingham-Elsevier Environment Corpus

Our corpus was compiled of papers from the journal Global Environmental Change (GEC) and from 10 other journals that relate either to the field of environmental science / environmental studies, or to disciplines that themselves are often included in environmental studies (life sciences, economics, social sciences). The journals were selected from those published by our partners in the project, Elsevier.[5] Our aim was to identify five monodisciplinary journals and five interdisciplinary ones, in addition to GEC.

Disciplinarity (that is, whether a journal should be classified as monodisciplinary or interdisciplinary) was assigned based on (i) normalized subject counts in Scopus[6] and (ii) the use of a clustering coefficient. To explain (i): if a journal belonged to a larger number of subjects than a typical journal in the same field, it was considered interdisciplinary, while if it belonged to a smaller number of subjects, it was considered monodisciplinary. With respect to (ii): Elsevier creates a map of journals based on citation relationships in Scopus, such that journals not citing each other are placed further apart than those that do cite each other. On the assumption that papers tend to cite papers in the same discipline, monodisciplinary journals should have most of the citation links nearby. Clustering coefficients reveal how well-connected a journal is in the map and take into account the connections of the other journals to which it is connected. Journals were considered as monodisciplinary if the coefficients indicate that they are connected to the journals that are well-connected to one another, and interdisciplinary if they are connected to journals that are not connected to one another. Of the monodisciplinary journals thus identified, those that mention interdisciplinarity in their aims and scope were excluded. This selection procedure resulted in the following journals being selected for the corpus alongside GEC: the interdisciplinary

[5] We are grateful to the research department of Elsevier publishers, who helped in the planning and execution of the project, including providing access to all the journals used in our corpus.

[6] http://www.scopus.com/

journals comprise Agriculture, Ecosystems & Environment (AEE), BioSystems (B), Computers, Environment and Urban Systems (CEUS), Environmental Pollution (EP), and the Journal of Rural Studies (JRS); the monodisciplinary journals comprise Advances in Water Resources (AWR), Journal of Strategic Information Systems (JSIS), Plant Science (PS), Resource and Energy Economics (REE), and Transportation Research Part D: Transport and Environment (TRTE). GEC has been in publication since 1990, and the other journals have been published since 1996 or earlier.

From GEC, the corpus includes all the articles from the first volume (1990/1991) up to Volume 20 (2010), while from the other journals it includes all the articles from the issues published between 2001 and 2010. The corpus includes full-length articles and excludes non-research papers such as book reviews. Only the main body text is included; other sections of the research papers such as abstract, footnotes, appendices, tables and figures are excluded. Since mathematical symbols and equations can cause problems in automated feature extraction, they have been replaced with the non-word *EQSYM*. For the sake of reliability in computing the frequency of linguistic features, it was necessary to exclude those papers whose body sections are 2000 words or less (Biber, 1990: 261). As a result, 501 papers (4.3%) were excluded.

Table 1 shows the size of the corpus, including the number of papers, the number of tokens, and the average length of paper in each journal in each year. In total, the corpus comprises 11,201 papers with a total corpus size of 51.4 million tokens. It is noticeable that the amount of data varies across journals. The EP texts, for instance, amount to 13.5 million words, whereas the JSIS and REE sub-corpora consist of only 1.3 million words each. Diachronic trends can also be observed. There tend to be more papers and a larger number of words in the latter half of 2000s than in the earlier years. Further, the average length of a paper differs between journals. While the mean length of texts in EP and PS is fewer than 4,000 words, the mean length in JRS is over 8,000 words.

3. The Dimensions

3.1 *Identifying and interpreting dimensions*

To perform multidimensional analysis, we first identified the linguistic features to be included in the analysis. The goal here is to include as many potentially important features as possible (Biber, 1985, 1988, 1995; Conrad & Biber, 2001). The present study started with the full list of over 150 features identified by the Biber tagger (e.g. Biber, Johansson, Leech, Conrad & Finegan 1999). Many were eliminated because they were at a high level of generality and overlapped with the more specific features that were retained. Others were merged to avoid redundancy. A few were eliminated or merged because they did not distinguish between the texts in our corpus. The result of this elimination and merging of categories was a list of 53 features; these are shown in Table 2. Each linguistic feature was identified and its frequency counted with the Biber tagger (Biber, 1988); this tagger has been developed over the past 30 years and extensively used in previous MDA research. The tagger identifies some features (e.g. demonstrative pronouns) with the aid of part-of-speech tagging,

and others (e.g. abstract nouns) by using vocabulary lists. Table 2 shows the normalized frequency (per 1,000 words) and the standard deviation of each feature.

For the present study, Professor Douglas Biber employed an exploratory factor analysis in order to identify co-occurrence patterns among the 53 linguistic features. A factor analysis of this kind reduces the number of original variables to a much smaller number of underlying variables, or factors. This process works by identifying systematic patterns of shared variance. The frequency of each linguistic feature varies across texts, so that Feature X may have a high frequency in a certain text but a low frequency in another. This pattern, or variance, is more or less shared by other features. If Feature Y shows a similar pattern to Feature X, the two features have a large shared variance.

This factor analysis was run on the normalized frequency of the linguistic features. Factors were extracted with principal factor solution, in which the first factor captures the largest amount of shared variance, the second factor captures the largest shared variance after the first factor is extracted, and so on. Multidimensional analysis requires the researcher to select the number of factors to be used. We decided on a six-factor solution based on the scree plot that shows the amount of explained variance, communality values that indicate the extent to which the variance of each feature is captured by the factors, and the factorial structure and its interpretability. A Promax rotation was applied to facilitate the interpretation of each factor. Table 3 shows the inter-factor correlation.

Table 4 lists the factor loadings for each feature in each factor. Factor loadings are the correlation between each feature and the factor and indicate the degree to which a feature is representative of the factor. For example, the factor loading of ‘abstract noun’ (shown in bold) for Factor 1 is 0.729, which shows that abstract nouns and Factor 1 co-vary closely. On the other hand, the factor loading of ‘nominalization’ (shown in bold) for Factor 6 is -0.789, which means that nominalization and Factor 6 demonstrate complementary patterns and that nominalization tends to be less frequent in the texts where the features with positive loadings (common nouns) are frequent. The factor loading of 0.30 was set as the cut-off point and a feature was retained in the final factorial model only if its loading was above the threshold. In common with standard practice in MDA, if a feature loaded over 0.30 in more than one factor, it was retained only in the factor on which it loaded highest. This practice ensures that the factors are entirely discrete.

Table 5 shows the resulting factorial structure. The factors were then interpreted as dimensions, using as input for the interpretation both the positively and, where relevant, negatively loaded features in each factor and close reading of extracts from the papers that are highly positive or negative on each factor. The dimensions will be presented here only in outline, for reasons of space. Exemplification will be reserved for the next stage of analysis, the Text Constellations. In each case our interpretative mnemonic for the dimension will be given together with a summary of significant features.

Dimension 1: system-oriented vs action-oriented

Factor 1 includes the largest number of features. Past tense and perfect aspect verbs are negatively loaded; present tense is positively loaded, as are determiners and abstract nouns,

including stance nouns, process nouns and cognitive nouns. Personal pronouns and activity verbs are also positively loaded. Our interpretation is that low-scoring papers are oriented toward actions (what the researcher did at particular times) whereas high-scoring papers are oriented away from action and time and towards a description of systems. The labels 'system-oriented' and 'action-oriented' capture the difference between papers that are not time-specific and are about abstractions and ideas rather than about actions and those that are clearly located in time and describe an experimental method.

Dimension 2: explicit vs implicit argumentation

Passives are negatively loaded in this dimension. Positively loaded features include modals (possibility, prediction and necessity) and adverbs of various kinds, including stance adverbs and conjuncts. Papers that score highly in this dimension are notable for the degree of explicitness of explanation. For example, clauses may be explicitly linked using conjunctive adverbs. The labels 'explicit argumentation' and 'implicit argumentation' capture the concern, or lack of it, for the reader's engagement with the text.

Dimension 3: spoken-ness

The features with a high loading in this factor are associated with spoken, conversational English. These include contractions, *that*-deletion and personal pronouns. There are no negatively-loaded features. The journals (JRS and JSIS) that have a relatively high mean score on this dimension (see Table 6) feature papers that report surveys and interviews with individual members of the public. Such verbatim reports account for the presence of spoken features in the dimension.

Dimension 4: conceptual discourse

This dimension is characterised by positively loading features only: word length, attributive adjective, coordinating conjunctions (connecting phrases), topical adjectives, and to-complement clauses that are controlled by stance nouns. Higher average word length in a text indicates greater informational density (Biber, 1988). Attributive adjectives provide conceptual elaboration and topical adjectives (which are also predominantly attributive) are used for classifying concepts (examples include *political, public, social, human, national*). Phrase connectors are used by writers to list, qualify, compare and contrast.

Dimension 5: text-focused vs site-focused

The positively loaded features in this factor are all connected with reported discourse and suggest a plurality of voices, or multiglossia. Papers with a high positive score on this dimension incorporate a variety of voiced opinions in their exposition and are thus focused on other texts. There is only one negatively loaded feature, place nouns. Our corpus contains a substantial number of papers that describe events and situations in specific geographical locales. Papers with a focus on places rather than on text have a high negative score on this dimension.

Dimension 6: non-research world vs research world

Factor 6 is somewhat curious, in that it consists of only two features, one positively and one negatively loaded. These appear to form a binary distinction, between common nouns and nominalisations. It is known that nominalisations are a feature of academic discourse, and it

is thought that they become relatively more frequent as fields of study progress and gain maturity (Halliday and Martin 1993). Although there is some difficulty in aligning Halliday's view of nominalisation with the 'nominalization' tag,[7] a study of relevant papers confirms a distinction between entities existing independent of the research process and those construed by that process. Although we recognise the anomaly of labelling any research article with 'non-research world', the two dimension labels provide a useful shorthand.

3.2 *Dimension scores and journals*

As noted in the Introduction, we first applied the dimensions to the corpus in a conventional way, by comparing the dimensional profiles of the various sub-corpora. In this study, the papers from each of the 11 journals comprised a sub-corpus. Comparing the dimension scores of the journals stands as a test of the validity of the dimensions; it allows us also to observe the 'style' of each journal, as well as to carry out further studies such as testing the degree of heterogeneity in each journal and to compare monodisciplinary and interdisciplinary journals.

The first step in this part of the study is to calculate where each paper is located along each dimension. For this purpose, dimension scores were calculated, representing the saliency of each paper in each dimension. A high dimension score shows that the paper has relatively high frequency of the positive features included in the dimension. Though dimension scores are based on frequency counts of each feature in each paper, the frequency cannot be used by itself for the computation because the absolute frequency of features varies widely (e.g. 'nouns' are very frequent in all the texts whereas 'relative clauses' are infrequent even in those texts where they are relatively frequent). In order for the features to be comparable, the normalized frequency of each feature was first standardized to a z-score, a value with the mean of zero and the standard deviation of one. After computing z-scores, dimension scores for each paper were calculated by summing the z-scores of the positive features in the dimension and subtracting the z-scores of the negative features. The calculation included only the features listed in Table 5.

The next step is to compute the mean dimension score for each journal. By comparing these scores, we can reveal the differences between the journals along each dimension. Figure 1 shows the mean dimension score and the standard deviation of each journal in each dimension, and Table 6 presents the results of analysis of variance (ANOVA) and post-hoc Tukey HSD tests. Since in all the dimensions the vast majority of journal pairs turned out to be significantly different in their mean dimension scores, the table lists the pairs whose mean dimension scores were NOT significantly different from each other.

There are 11 journals, meaning that there are 110 pairs of journals altogether. If a dimension was completely unsuccessful at distinguishing between journals, a list of close to 110 pairs would appear under that dimension in Table 6. If a dimension was completely successful, then no journal pairs would be listed under that dimension. Table 6 shows that neither is the case, but also that, given the large number of total possible pairs, all the dimensions are

[7] Nouns tagged by the Biber tagger as 'nominalizations' are identified only by suffixes such as '-tion'. Whereas these do identify genuine nominalisations, there are some false hits and not all nominalisations are identified by this method.

relatively successful. According to the table, dimension 1 is the most successful in distinguishing between journals, as only two pairs are not distinguished by the dimension. Dimensions 2, 3 and 6 are the next successful. Dimension 4 is the least successful, with 9 pairs of journals failing to show significant difference and only one journal (REE) not appearing in any of the pairings. If we follow the fortunes of one journal, Global Environmental Change (GEC), it is similar to TRTE in dimensions 1 and 5, to JRS in dimensions 2, 4 and 5, to JSIS in dimension 2, to AWR in dimensions 5 and 6, and to CEUS in dimension 6. This suggests that dimension 5 is the least successful in distinguishing this journal from the others and that dimensions 3 and 4 are the most successful. We might expect that interdisciplinary journals would be less distinctive than monodisciplinary ones. Counting how many pairs each journal enters into, shown in Table 7, there is some variation, from TRTE showing a lack of distinction in 11 pairs and PS, B and REE showing a lack of distinction in only 2 pairs each. There is, however, no clear division between monodisciplinary and interdisciplinary journals, as Table 7 shows.

To summarise so far: the six identified dimensions are reasonably successful in distinguishing between the eleven journals, suggesting that they do capture distinctive features of research discourse style. Four of the journals included in the study are more distinctive in style than the others. However, the MDA methodology does not distinguish between monodisciplinary and interdisciplinary journals. This suggests that there is nothing that is stylistically distinctive between journals so identified.

The research questions in the Introduction ask also about the degree of heterogeneity within each journal, that is, the extent to which papers in a journal share similar patterns of the dimensions of variation identified through MDA. An additional procedure was carried out to determine this, based on classification and clustering analysis. A machine learning algorithm called random forests (Breiman, 2001) was employed to predict the journal of each paper based on the dimension profile of the paper. If the model can accurately classify a paper into the journal it was taken from, it suggests that the paper has a similar dimension profile to that of the journal as a whole. The random forests algorithm first produces a large number of tree-type classifiers on bootstrap samples (i.e. a randomly selected sample of original data) by using a subset of predictors. Each tree models the relationship between dimension profiles of individual papers and the corresponding journals. Random forests then uses the models to generate prediction (Kuhn & Johnson, 2013). With respect to clustering, a hierarchical agglomerative cluster analysis was run on 11,201 papers with dimension scores as the features. We used Euclidian distance with the Ward agglomeration method in clustering.

The random forests showed that the total out-of-bag prediction accuracy was 71.9%. This means that in total 71.9% of all the papers were accurately classified into the journals they were published in. This is significantly above chance ($\chi^2(1) = 3823.42$, $p < .001$, $\phi = 0.413$), with the chance being the probability that all the papers are categorized into the largest category (EP; 30.6%). The value of 71.9% is high, considering that the algorithm had 11 journals to choose from. Thus, we can conclude that the distinct dimensional profile of each journal generally applies to individual papers.

This, however, does not apply to all the journals equally. Table 8 is the confusion matrix of the random forests classification and shows the number and the proportion of the papers in each journal that are classified in each journal. For instance, the top left cell shows that 999 papers, or 58.1% of the papers, in AEE were accurately classified as AEE papers, while the next cell to the right shows that 40 AEE papers (2.3%) were wrongly classified as AWR papers. We can tell from the table that some journals, such as AWR, B, and EP, have high classification accuracy (> 80%), whereas others, such as CEUS, REE, and TRTE, have low accuracy (< 40%). This suggests that some journals have their own unique writing styles while others publish papers of varying writing styles.

Figure 2 shows the results of clustering. There are 11,201 rows in the figure, each corresponding to a paper. The leftmost panel shows the journal the paper was published in. The middle heatmap displays the standardized dimension score of each paper. Darker shades show that the dimension score of the paper was high, while lighter shades (or white parts) indicate low dimension scores. The rightmost panel is the dendrogram of the cluster analysis. It appears that the three-cluster solution is appropriate as there is relatively a large gap between the second and the third division from the root. Assuming three clusters, dashed lines were drawn between clusters. Based on the figure, some observations can be made. First, the distinct profile of dimension scores in each cluster can be seen, and the distinctiveness is particularly brought by Dimension 1 and Dimension 6. Papers in the top cluster tend to have relatively high dimension scores in Dimension 1 and low dimension scores on Dimension 6, while those in the middle cluster tend to score low in both dimensions. Papers in the final cluster tend to score high in Dimension 6.

Second, papers in each journal do not equally spread across the three clusters, but there is some variation in how spread the papers are depending on journals. Papers in B and JSIS, for instance, are almost exclusively clustered into the final cluster, while the vast majority of the papers in CEUS, AWR, GEC, REE, and TRTE are clustered in the top cluster. Papers in the other journals more or less cut across clusters. EP papers spread over the first two clusters, while JRS papers are clustered both in the top cluster and the final cluster. PS papers are densest in the middle cluster, but a nontrivial number of their papers are also clustered in the other two clusters. AEE is interesting in that, although there appear to be the largest number of papers in the final cluster, a significant number of papers were clustered in the other two clusters as well. This implies that AEE papers vary with regard to their writing styles, and indeed, the random forests discussed above shows relatively low classification accuracy (58.1%), suggesting that there is much intra-journal variability.

In order to test the magnitude of variation within and between journals, the dimension score was z-transformed within each dimension, and a multiple regression model was constructed that models the standardized dimension score of each paper in each dimension as a function of dimension, the journal the paper was taken from, and their interaction. The results showed that the model explains 43.8% of the variance ($F(65, 67140) = 807.4$; $p < 0.001$; adjusted $R^2 = 0.438$). This suggests that more than half of the total variance is attributed to within-journal variation. Thus, while journals distinguish dimension scores to a certain extent, there is considerable variation within each journal as well.

4. Text Constellations: multi-dimensional analysis from the ground up

4.1 *Identifying constellations*

Our main aim in carrying out the multidimensional analysis was to find a bottom-up way of identifying sub-corpora within the target journal Global Environmental Change and so to establish the degree and the nature of diversity within that single interdisciplinary journal.

To do this, we clustered individual GEC papers into the groups that share similar patterns of dimension scores (cf. Biber, 1989). More specifically, we first z-transformed the dimension scores of GEC papers within each dimension so that the scores were comparable across dimensions. We then ran a hierarchical agglomerative cluster analysis with squared Euclidean distance and the Ward clustering method. The resulting dendrogram (Figure 3) suggests that a range of numbers of clusters could be supported. Three is the most obviously optimum number, but we wished to have, in the initial stages, a more fine-grained and therefore informative set of clusters, and so selected six for the initial investigation. Each cluster corresponds to a number of papers in GEC that are similar in terms of their dimensional profile. To avoid a confusion of terminology, we have appropriated the term 'Text Constellation' to refer to the groups or clusters of papers thus identified. There are 118 papers in Constellation 1, 169 in Constellation 2, 61 in Constellation 3, 95 in Constellation 4, 35 in Constellation 5 and 146 in Constellation 6 (see Figure 4). Each paper in the corpus has been annotated in SketchEngine with its constellation number, allowing us to investigate the constellations as sub-corpora. Figure 3 also shows how the constellations relate to each other. The most distinctive constellation, with the highest tree-branching, is constellation 1. The constellations with the most similarity are constellations 3 and 6, and constellations 5 and 4. These four constellations together are distinguished from constellation 2.

Figure 4 shows how the dimensions map on to the identified constellations. Since the values are standardized, zero represents the grand average and is indicated by dashed lines in the figure. From Figure 4 we can see that constellation 2 has values that are closest to average across four of the six dimensions, while constellations 3 and 5 have values that diverge considerably from the average. Constellation 3 has a fairly narrow range of values along each dimension, suggesting papers that are relatively homogenous, while constellation 5 has a broader spread in at least three dimensions, suggesting greater variety between papers. Some constellations with different average scores in one dimension nonetheless show overlap, such as constellations 4 and 6 with respect to dimension 6. What Figure 4 enables us to do is to identify the dimension features of each constellation, and also to visualise the degree of difference between the constellations.

Although diachronic comparisons are not the focus of this paper, it might be noted in passing that it is possible to track the progress of the constellations through the 20 years of the GEC corpus. Figure 5 shows this via a 'moving window': each point on the x axis represents three years rather than one, and each point moves on one year from the previous one (e.g. 1999-2001, 2000-2002, 2001-2003 and so on). The y axis represents the proportion of all the papers in each three window comprised of each constellation. The figure suggests that,

whereas each constellation has peaks and troughs of popularity, the relative proportion of the journal occupied by each constellation has not altered significantly over time; however the marked divergence observable in the first three years (constellation 2 overwhelmingly frequent, constellation 5 barely figuring) has given way to a relatively more even spread in the final three years.

Our method, then, has reversed the usual MDA process, deriving the Text Constellations by comparing the dimensional profiles of individual texts. Each of the constellations should represent a distinctive type of paper in GEC. In practice, because of the varying degrees of similarity and overlap, it is easiest to demonstrate difference between the most widely distinguished constellations and then to describe the others in relation to them. For this reason we begin with a detailed description of constellations 1, 5 and 3. Figure 4 shows constellations 1 and 5 as being visually the most different from each other and so the most easily distinguished and contrasted. Figure 3 shows 5 and 4 as being similar to each other, and 3 and 6. For this reason, constellation 3 is added to this initial description as representing that pair of constellations. We describe these constellations first via the dimension profiles in 4.2 below, and then via phraseological evidence in 4.3. We then proceed to discuss the other constellations (4.4).

4.2 *Constellations 1, 5 and 3: dimensional profiles*

In describing these three constellations we rely for the most part on the features occurring on the relevant dimensions. For example, references to 'degree of 'spoken-ness'' reflect an interpretation of the positively weighted features on dimension 3 (contractions, use of pronouns, *that*-deletion and so on). In some cases, however, scrutiny of many papers in a constellation has led to the identification of other recurring foci that do not appear in any of the dimensions. For example, many of the papers in constellation 5 articulate an antagonistic stance towards a purely 'scientific' approach to studying environmental change. Examples are given below. Similarly, many of the papers in constellation 1 express concern and pessimism about the likely future environmental changes, and their consequences, in the locations they have studied. Again, examples are given below. In neither case is this attitudinal information retrievable from the factor/dimension loadings. Indeed, they are not and could not be identified and tagged in the corpus from which the factor loadings are calculated. They do, however, represent distinctive features of the constellations in question, occurring as corollaries to the tagged features. The examples given below are taken from a broad diachronic range. These examples are inevitably selective. The aim is not to demonstrate typicality via the examples; typicality or representativeness of the features identified has been demonstrated through the multidimensional analysis and consequent identification of the constellations. For example, in constellation 5 there is an engagement with other research voices, identified through the relatively high incidence of communication verbs in dimension 5. This is illustrated with some examples of engagement with other researchers. The absence of such examples from the account of constellation 3 does not mean that papers in that constellation do not cite other work, but rather that the factor analysis has indicated that communication verbs are less salient in constellation 3 than in constellation 5.

A dimension-led description of each constellation follows.

Constellation 1: site- or target- specific narrative and quantification

This constellation scores low on dimensions 1, 2, 3 and 5 and high on dimension 6 (see Figure 4). This suggests: a concern with action or events rather than with system; use of implicit rather than explicit argumentation; a relative absence of features associated with spoken-ness; a focus on space and place rather than on text; and a concern with the non-research world. This might be summarised as a relative prioritisation of the physical world over the world of ideas, combined with a relative lack of concern to explain steps in argumentation to the reader. The papers belonging to this constellation tend to focus on specific sites of interaction between people and the environment (e.g. forest, coastal cities, individual countries or regions), often coupled with specific influences on environmental change. Most of the papers in the constellation give quantified data about changes in aspects of the environment and construe human societies as abstractions defined by environment-related activity. In spite of this apparently 'de-humanised' approach, most papers also attach value judgements to predictions about climate change and environmental loss.

Examples of the characteristics of constellation 1:

Focus on place: *Vegetation in the Great Basin prior to domestic grazing can be broadly discerned from the journals and diaries of early European-descent travellers through the Great Basin.* 1996_Knapp

Implicit argument: *The work described in this paper was conducted as part of a broader scenario analysis… The scenario analysis compared the energy, emissions, and economic implications…with a set of reference technology assumptions…These four sets of technology assumptions were compared in the context of…* 2008_Thomson

Focus on quantity: *...it involved the creation of five major reservoirs that have flooded 9675 km of boreal forest and two major river diversions totalling -1600m.3/sec, about twice the flow of water diverted out of the Churchill River.* 1995_Rosenberg

Abstracted human action: *A growing urban and middle-class segment of the national population could also mean changing perceptions of the forest.* 1999_Mather

Value judgements and predictions: *At present, however, it seems only too likely that by very soon after 2000 all but the most inaccessible parts, and a few reserves, of the rich forest environment that endured the ice ages and 10,000 years of subsequent human history will have been irreparably destroyed for gain in the space of just half of one century.* 1990_Brookfield

Constellation 5: personal voices

This constellation is distinctive in that the spread of scores between the dimensions is greater than those of the other constellations. It scores high on dimensions 1, 2, 3, and 5 and low on dimensions 4 and 6. This suggests: a focus on system rather than action; a concern to make

arguments explicit; a relatively high proportion of features associated with spoken-ness and with a text focus; a concern with the research world. In summary, there is a focus on the abstract but also on engagement with a number of voices and with explicit argumentation. The papers belonging to this constellation deal with human perspectives on the environment, including perception studies, and also with social perspectives of science. This is a smaller constellation than constellation 1, with only 45 papers.

Examples of the characteristics of constellation 5:

Focus on system: *We have argued that problem definition depends upon interest and perspective.* 1995_Herrick
Vulnerability arises through particular levels of exposure to underlying socio-economic changes and to climate-related impacts flowing from the different scenarios. 2000_Lorenzoni

Person-centred methods: *I will demonstrate this with an analysis of long interview discussions that I staged in Bristol, UK in 1992 and 1993.* 1996_Hinchliffe
Of the physicist trio, two were interviewed in person, and showed themselves to be remarkably frank. It was not possible to interview NAME, wherefore I resorted to numerous persons who knew him. 2008_Lahsen

Construal of a negative orientation to science: *The responses were seriously at variance with the scientific models of global warming.* 1991_Kempton
However, a growing area of scholarship stresses the need to also study the role of culture and politics in the very production of scientific knowledge and associated adjudications (references). 2008_Lahsen

Spoken-ness: *Put simply, how do institutions constrain and shape behaviour when they are themselves the products of human choices?* 1999_O'Riordan
Ok, this object is cool – actually it is toxic when burned, but we don't care anyway, as we don't exactly know what effects we cause. 2001_Stoll-Kleeman

Engagement with research voices: *At the core of this is what Graftstein (1992) terms the 'paradox of constraint'.* 1999_O'Riordan
Building on Amartya Sen's entitlement approach (reference) and other sources, asset-based and livelihoods approaches state that household well-being is multi-dimensional…(reference). 2009_Heltberg

Explicit argumentation: *Because empirical concepts are open textured, a science-based assessment of a policy related issue can always be charge with 'sins of omission'. For instance, acquatic [sic] damage from acid deposition can be characterized in several ways. If damages are stated in terms of the number of lakes affected, then projections of decreased deposition appear to provide a substantial decrease in damages….Moreover, the choice of a reference pH value can radically alter the number of 'acidic' surface waters… Still another consideration…* 1995_Herrick

Constellation 3: modelling

This constellation scores high on dimensions 1 and 6 and relatively low on dimensions 2, 3 and 4. This suggests: a concern with system rather than action but a complementary concern with the non-research world; relatively little use of explicit argumentation; little use of features connected with spoken-ness or conceptual discourse; relatively little text focus. It contrasts with constellation 1 mainly in respect to dimension 1. In other words, constellation 3 is more system-oriented whereas constellation 1 is more action-oriented. It contrasts with constellation 5 with respect to dimensions 3 and 5 in particular, implying that has fewer features associated with spoken-ness and with textual interaction. The papers in this constellation are mostly about the activity of modelling environment change.

Below are two examples of constellation 3 from the two decades of the GEC corpus, demonstrating the interactions between the physical world (e.g. *atmospheric processes; carbon emissions*) and the world of mathematical projection (e.g. *our ability to predict accurately; two alternatives…system*).

Focus on system: *What should be recognized is that the major sources of greenhouse-gas emissions are known and that the role of these gases in influencing climate is well understood and accepted. The controversy surrounding the general global warming which may accompany the rapid increase in greenhouse-gas emissions is based on our ability to predict accurately the effect these gases will have on climate in the presence of other atmospheric processes.* 1991_Lonergan, emphasis added

Focus on the non-research world: *We discuss two alternatives for a domestic system of carbon emissions trading. Option I caps carbon at the point of production. A system capping carbon at the producer is termed an "upstream", "supply" or a "fuels" system. Under option I, we specify that permits are auctioned. Option II is a "downstream", "combustor", or "end-user" system that controls carbon at the point of fuel combustion. For this downstream design, most of the compliance permits are allocated free to controlled entities*. 2000_Holmes, emphasis added

Examples of this constellation also illustrate the focus on uncertainty and imprecision that makes modelling a recursive activity (e.g. *estimating, complex task, results vary, uncertainty, assume, overestimated, underestimation* in the example below):

Estimating the amount of GHG emitted from deforestation is a complex task and the results vary considerably. Both deforestation rates and the emissions per deforested hectare (EpH) are subject to uncertainty. … Here we will assume that the conversion of one hectare emits the amount of carbon stored in its above and below ground biomass. By ignoring the fraction of carbon that can be stored in the subsequent land cover, the fraction not immediately released into the atmosphere or carbon trapped in long-term wood products, emissions will be overestimated. On the other hand, ignoring carbon emissions from soil and other GHG emissions leads to an underestimation. 2009_Strassburg, emphasis added

The examples and descriptions above suggest that the three constellations do indeed occupy

different spaces in the overall research paradigm. Constellation 1 is the most 'science like', reporting empirical work. Constellation 5 is the most 'social science like', reporting social and political attitudes and responses to environmental change. Constellation 3 is the most mathematical and in some cases articulates a mixed method of working.

4.3 *Constellations 1, 5 and 3: phraseology*

Having identified the sub-corpora comprising constellations 1, 5 and 3 using quantitative methods, the distinctions between them might be explored in a more qualitative way by investigating the phraseology of selected words. In order to do this, as noted above, each paper was tagged with its constellation ID number and the SketchEngine interface was configured so that each constellation could be accessed and searched independently. In this section we briefly consider the results of a small sample of investigations of individual words and phrases, to illustrate a more general point about interdisciplinary discourse rather than to provide a comprehensive account of the phraseology of the journal.

The first word to be studied in this way is *environment*. The collocates of *environment* in each of constellations 1, 5 and 3 are identified, using a span of ±3, and ordered using the logDice statistic in SketchEngine. The 20 most significant collocates in each constellation are as follows:

Constellation 1: ***terrestrial***, ***biophysical***, *enabling,* ***semi-arid***, *Hydrometeorology, Development,* ***physical***, *Conference, Nations, Agency, Terrestrial, Ministry,* ***natural***, *Stockholm, spatially, Middle,* ***marine***, ***forest-savanna***, *Improvement,* ***oceanic***.

Constellation 5: *Agency, DEFRA, Department, sake, business, protecting, bad, campaign, evaluated, care, protect, Programme, you, responsibility, Nations, Earth,* ***natural***, *United, towards, concern*.

Constellation 3: *Outlook, Agency, Fisheries, Rural, Department, Ministry, man, Terrestrial, economy, Minister, European, Global, Conference, UNEP, Development, Nations, attitudes,* ***built***, *Canada, value*

Most of the items in all the lists, and especially in constellation 3, comprise names of organisations and events. Other than that, the lists are different, not only in the actual words (types) found, but in the classes of types and the consequent implied construal of *environment* as an entity. Items shown in bold italic above provide a classification system for environments based on physical characteristics. There are 8 such items in the constellation 1 list (*terrestrial, biophysical* and so on), but only one (*natural*) in the constellation 5 list and one (*built*) in the constellation 3 list. Items shown in single underlined italic above construe the environment as a vulnerable entity requiring care and protection, an entity attracting moral duty (*responsibility towards*) and an affective response (*concern*) on the part of the public. These all occur in the constellation 5 list. Thus the 'physical science' / 'social science' divide between constellations 1 and 5, noted in the previous section, is confirmed. Constellation 3 is different from both, and with only five items that are not Proper Nouns is

more difficult to interpret. The words *man* [in context meaning 'human', sic] and *attitudes* suggest a human focus, while *economy* and *value* suggest an economic focus.

Further confirmation can be found by looking at 3-grams. Lists of n-grams (2-, 3-, 4-, and 5-grams) in each of the six constellations have been compiled. Unsurprisingly, these show the 3-grams *of climate change* and/or *to climate change* occurring close to the top of the relevant frequency lists in all constellations. There are, however, some notable differences. Again focusing on constellations 1 and 5, it is perhaps surprising that whereas in constellation 5, *of climate change* and *to climate change* appear as 2nd and 3rd respectively in the 3-gram list (ordered by raw frequency), in constellation 1, *of climate change* appears 5th in the list but *to climate change* only 53rd. The raw frequency 'gap' between the two phrases diverges too. In constellation 5, *of climate change* occurs 223 times and *to climate change* 166 times; in constellation 1, the respective figures are 303 and 97. This might be summarised as: *of climate change* is an important phrase in both constellations, but *to climate change* is more important to constellation 5 ('social science') than to constellation 1 ('physical science'). Constellation 3 is very similar to constellation 5, with *of climate change* and *to climate change* appearing 2nd and 4th respectively in the 3-gram list, but the gap between them is greater, with frequencies of 382 and 193 respectively.

It comes as no surprise, then, that a more in-depth investigation of the phrases reveals further differences. The choice of preposition (*of* or *to*) is of course dependent on the word that precedes it (e.g. *consequences of* but *adaptation to*). The lists below show all the L1 collocates of *of climate change* occurring with a frequency of 3 or more in each of constellations 1, 5 and 3.

Constellation 1: *aspects, assessment, because, consequences, effect, effects, estimates, forces, impact, impacts, implications, magnitude, patterns, projections, rate, result, scenarios, studies*.

Constellation 5: *aspects, because, context, danger/s, effects, era, impact, impacts, implications, issue, perceptions, problem, risks, source/s, threat, understanding*.

Constellation 3: *absence, analysis/es, aspects, assessment, because, consequence, consequences, cost, damages, effect, effects, estimates, impact, impacts, implications, level, levels, magnitude, mitigation, perceptions, result, risks, scenarios, signal, study/ies, threat*

These lists are different in non-random ways that align with the known differences between the constellations. In constellation 1 there are words (*magnitude, rate*) relating to quantity. In constellation 5 there are words (*danger, effects, issue, problem, risks, threat*) relating to danger. In constellation 1, nominalisations (*assessment, estimates, projection, studies*) construe research processes. In constellation 5, the equivalent nouns (*perceptions, understanding*) construe everyday mental processes. The constellation 3 list includes nouns that also occur in the constellation 1 list (*assessment, estimates, magnitude, scenarios, study/ies*), some that also occur in the constellation 5 list (*impact/s, implications, perceptions, risks, threat*) and some that are unique to this constellation (*absence, cost, damages, level/s, mitigation, signal*). Thus constellation 3 is concerned with quantifying climate change and its

effects (*assessment, cost, level, magnitude*) and also with hypothesising (*in the absence of climate change, analysis, scenario, risks*), as well as perceiving climate change as danger (*damages, risks, threat*).

The 3-gram *to climate change* shows much less divergence in terms of collocation, even though its relative frequency shows a difference between the constellations. The list of L1 collocates occurring three or more times is longer for constellations 3 and 5 than for constellation 1 but not otherwise markedly different:

Constellation 1: *adapt/adaptation, due, related/relating, respond/responses, sensitive/sensitivity, vulnerability/vulnerable*.

Constellation 5: *adapt/adaptation, approach/es, contribute/contribution, regard, related/relating/relation, resilient/resilience, respond/responses, vulnerability/vulnerable.*

Constellation 3: *adapt, adaptation, adapting, contribution, contributions, due, relation, response, responses, sensitive, vulnerability, vulnerable*.

However, a study of the concordance lines for *responses to climate change*, relatively frequent in all three constellations, shows the expected difference. In constellation 1, agents of the responses are organisations (*CERES' responses to climate change inputs*), the natural world (*the differential ecological response to climate change; potential adaptive responses to climate change; vegetation responses to climate change*) and abstractions of human groups (*the drivers of agricultural responses to climate change; concerned with poverty responses to climate change*). In constellation 5, the agents are all human, construed either as individuals (*understandings of and responses to climate change; the diversity of responses to climate change…throughout our discussions; shaping people's responses to climate change*) or as groups and abstractions (*community responses to climate change; international responses to climate change; migration responses to climate change; developing sustainable responses to climate change*). In constellation 3, where *response* (singular) *to climate change* was investigated because it is more frequent, the agents of response are sometimes the natural world (*vegetation and water-cycle responses to climate change; the ecosystem response to climate change*) but more often human agencies (*the societal response to climate change; the domestic response to climate change; the economic response to climate change; if water supply is augmented in response to climate change*). What is striking about the *response to climate change* examples in constellation 3, though, is that modelling is consistently mentioned: *the reverse arrow indicates…; modelled in CLIMAPS; …a separate sub-model of IMAGE; …non-linear response; …two climate models*. In other words, the responses are not actual ones but modelled ones.

4.4 *The other three constellations*

As noted above, the greatest similarity measures shown in figure 3 are for constellations 3 and 6, and for constellations 4 and 5. In this section, constellation 6 will be described in comparison with 3, constellation 4 in comparison with 5, and constellation 2 will also be described. These descriptions are carried out in relation to the dimension profiles shown in

Figure 4.

Constellation 6 ('modelling human beings') is similar to constellation 3 in terms of dimensions 1 (high in both) and 4 (low in both). It is also similar to constellation 5 in terms of dimension 2 (high in both). It contrasts with constellation 1 in all dimensions except 4. The dimensional profile suggests: a focus on system rather than action; a concern for explicit argumentation; a focus on text but not on conceptual discourse. Like constellation 3, these papers explore models and uncertainty, but the models have a more human focus, as shown in these examples:

The distinction between uncertainty and indeterminacy is important because the former enshrines the notion that inadequate control of environmental risks is due only to inadequate scientific knowledge, and exclusive attention is focused on intensifying that knowledge, to render it more precise. Very often this extra technical precision is a surrogate for more 'precise' control of social actors and the indeterminacies they bear. 1992_Wynne

Imagine a set of actors, each owning definite quantities of various goods. These actors meet on a market place, where an auctioneer proposes an arbitrary price scheme for these goods. Each actor has a well-defined pattern of preferences for all the combinations of goods which they could possibly consume. 1996_Jaeger

Given the technical difficulties and expense of monitoring carbon stock changes at the farm level, incentives may need to be based on activities rather than upon measured changes in the soil. Decoupling incentives from carbon accounting would allow for incentive payments to focus on those practices that have the highest environmental benefits. 2000_Subak

Constellation 4 ('researching people'), as noted, is somewhat similar to constellation 5. Both constellations score relatively high on dimensions 3 and 5. Both score low on dimension 4. They are distinguished in respect of dimension 1(where constellation 4 scores low and constellation 5 scores high) and dimension 6 (where constellation 4 scores high and constellation 5 scores low). This suggests that constellation 4 will be more action-oriented than is constellation 5 (more concerned with the process of research itself), and more concerned with the world of things (common nouns) rather than abstractions (nominalisations). In practice, constellation 4, like constellation 5, focuses on people, and includes histories of academic and political approaches to issues of environmental change as well as surveys of public attitude.

The following two examples from this constellation illustrate the focus on action, and also the interaction with human subjects:

A panel of five expert climatologists was selected and assembled to develop future climate scenarios and their controls. The experts were individually asked to identify and explain each of the current (1993) climatic controls and to detail the anticipated changes in such controls which would produce climate change at the Yucca Mountain site. 1995_Miklas

Respondents were also asked to choose up to three actions, from a list, which they thought would best tackle climate change. 2008_Pidgeon

The next example illustrates reflexivity about discipline, and the combination of natural sciences and social sciences:

We are approaching our research on soil erosion at Fandou Béri from two directions, reflecting our disciplinary backgrounds as social and natural scientists. First, we are measuring the erosion itself. ... The household's history since 1960, income and expenditure patterns, labour patterns, and demographics of the members of each household have been researched…. 2001_Warren

The final example shows the reflection upon the history of research into environmental change:

The scientific community has repeatedly claimed that it will be able to provide more certainty in future in order to improve the rational basis for policy, but reveals ever more uncertainties as the timespan needed for reducing them, once proposed for the 1990s, now extends further into the next century. 1994_Boehmore-Christiansen

Finally, constellation 2 ('theory') is the largest constellation in the journal, with 169 papers. Possibly as a result of the size factor, most of the dimension scores are around the average, and slightly above the scores in constellation 1. However, the constellation scores relatively high on dimension 4 and low on dimension 6. This suggests that the constellation will be action-oriented but will also be discursive, with an emphasis on building an argument around other researchers' contributions. Some papers in this constellation construct a history of research into environmental change, others address theoretical stances taken by various schools of thought, while others conduct more traditional meta-analyses of existing data. Here are some examples:

Focus on action: *Just as SCOPE began its activities, UNESCO established its Man and the Biosphere (MAB) programme in 1971. Many of the objectives of the MAB programme are similar to those of SCOPE, and this initially led to tensions with respect to projects which could be undertaken within either programme. In some countries, one committee guided research within both programmes; other countries established a committee for each.* 1990_Price

Focus on research paradigms: *Proponents of the pluralist paradigm see increasing social differentiation as the central societal process. By this it is meant that the division of labor increases as industrialization proceeds and as society becomes more complex.* 1995_Sunderlin

Meta-analysis: *In a recent application of meta-analysis in the field of land-cover change, Geist and Lambin (2001), dissatisfied with the limitations of cross-national statistical analyses, examined 152 cases of tropical deforestation and found synergistic combinations of direct and indirect factors causing deforestation, amongst which economic, institutional and political factors were prominent.* 2005_Misselhorn

5. Divergence and convergence in Global Environmental Change

The identification of constellations as outlined above has indicated the nature and degree of divergence between disciplines or approaches within the journal Global Environmental Change. The recognition of this divergence is not surprising; indeed, a key aim of much research in corpus linguistics is the identification of contrast and divergence between corpora. Specifically, the methodology of MDA is designed to identify just such difference. It is important to note, however, that in this corpus there is convergence between the constellations as well as divergence. The paradox between convergence and divergence reveals itself in two ways. Firstly there is the fact that papers in GEC show similarity as well as distinction. The random forests investigation (Table 8) shows a relatively high degree of similarity among GEC papers. The constellation profiles (Figure 4) show overlap between them. There is evidence from the phraseological evidence too. To return to the example of the collocates of *environment*: in the discussion above, differences between constellation 1 (*terrestrial, biophysical, semi-arid* etc) and constellation 5 (*protecting, campaign, care, responsibility* etc) were highlighted. It is also true, however, that references to institutions and organisations occur frequently in all constellations. Table 9 shows the equivalent list of collocates for each constellation. The lists comprise the top 20 collocates of *environment*, in a ±3 span, ordered by logDice. As can be seen, each list is distinctive, but there is also considerable overlap. For example *Nations* occurs in all the constellation lists and *Development* and *Conference* occur in five of the lists. *Programme* occurs in four and *UNEP* in three. The words entirely in lower case are more revealing, as these are not parts of the names of institutions. The word *natural* occurs in the constellation 4 and 6 lists, as well as 1 and 5. *Mountain* occurs in the constellation 6 list as well as 1. *Protecting* occurs in the constellation 4 and 6 lists as well as 5.

To add more perspective to these observations, it is worth considering uses of the word *environment* that are not found in the GEC corpus. The following examples of *environment* have been selected from the British National Corpus and illustrate uses of the word ranging from the physical (*damp environment*) to the social (*a caring family environment*) and the emotive (*a happier environment*):

…an **academic environment** appropriate to degree-level work
In the cold and possibly **damp environment** of the loft…
…who flourish in a **para-military environment**
…working in a **polluted factory environment**…
…within a **caring family environment**
…a **happier environment** in which to work
…to achieve a **safe environment** for patients

Examples similar to these do exist in GEC. These were identified by a search of L1 collocates in *environment* in the corpus as a whole (that is, all the constellations together). Modifiers indicating non-physical meanings of *environment* are: *political, social, institutional, economic, regulatory, business, decision-making, high-risk*. There are a few instances of modifiers, such as *clean* and *healthy*, that refer to an individual's immediate surroundings

rather than to an eco-system. Such examples are rare, however. One paper from the early days of the journal attempts to constrain the meaning of the word, and thus the scope of the journal:

> *…a brief note is warranted about the meaning of 'environment'.* ***Some segments of the social sciences and humanities use the term to refer to the social environment or circumstances of human life.*** *In this use, environment includes … not only the material conditions external to the individual…but the non-material as well… This broad meaning of environment is not the one that prevails* ***in the natural sciences*** *and in* ***those social science fields closely linked to them****, where* ***the term refers exclusively to the physical conditions of Earth****… The study of global environmental change, including its human dimensions,* ***clearly employs this second meaning of environment****.* 1990_Turner (emphasis added)

This metalanguage, used explicitly to establish the field of the journal, performs functions that might be described as 'doing divergence' and 'doing convergence', and this performativity is the second way in which divergence and convergence is enacted. The writer establishes the GEC field as divergent from 'some segments of the social sciences and humanities' but also establishes it as a convergence of 'natural sciences' and [other] 'social sciences', based on their common understanding of the key term in their field. Although this lengthy demarcation of meaning is rare in the corpus, a distinguishing feature of GEC is the explicit attention paid to academic discipline. One example of this is the antagonism shown by some writers towards the pure physical sciences, which are presented as either limited or distorting:

Inability to cope adequately with mounting environmental problems raises questions about whether existing scientific knowledge is adequate to the tasks in hand, and whether it is being put to best use. 1990_Boxer

Scientific analysis moves along pathways that distort the character of the phenomenon being studied. 1991_O'Riordan

Attempts to negotiate both the difference between disciplinary paradigms and the common ground of the journal may be observed in other ways too. Many of the introductions to the papers, especially in the early days of the journal, begin with very general statements. Here are some examples from the 1990-91 volume:

As we enter the last decade of the twentieth century, new views of our world are evolving. 1990_Price

The long sweep of human history reveals an escalating trajectory of alterations and transformation of Earth – of the geosphere-biosphere that sustains life as we know it. 1990_Turner

Should climate change as most global climate models predict, society will be confronted with the unavoidable task of adapting to a wide range of environmental impacts that could be costly and disruptive. 1991_Meo

It might be said that these authors are reaching out beyond their own disciplines to engage as broad an audience as possible. In the later years of the journal, explicit statements of the relevance of new approaches can be observed. For example, a paper that investigates responses to environmental change by First Nation peoples of Canada begins by 'doing convergence', establishing common ground:

With growing global recognition of climate change as a real, ongoing and accelerating phenomenon, there is a need to understand what effects are anticipated, and how human societies may be able to adapt… 2009_Turner

It proceeds by 'doing divergence', proposing a novel way of approaching the problem:

Turning for help and insight to Indigenous Peoples makes great sense… 2009_Turner

Our proposal of the terms 'divergence' and 'convergence' to describe the discourse features of the interdisciplinary journal we are describing goes some way to answering the final discussion question posed in the Introduction: what are the implications for the practice of writing in interdisciplinary research? We would suggest that successful interdisciplinary projects are often devised to demonstrate their common ground (their convergence), for example the extent to which they address a common question, and the unique contribution brought by each of the disciplines (their divergence). It may be, though this remains a question for debate, that successful interdisciplinary journals maintain over years both their common focus and their willingness to allow into that focus new disciplinary approaches. For example, in GEC, another paper from the later years of our corpus begins with a general statement:

Climate change is a complex problem that demands collective solutions 2009_Nerlich

and proposes research into language as a contributor to these solutions:

...attention should be paid to collective processes of sense-making ... such as the repeated and prolific use of compounding... 2009_Nerlich

In short, the notions of convergence and divergence may be useful in conceptualising interdisciplinary research, and its discourse, more generally.

6. Conclusion

This paper has demonstrated a novel application of multidimensional analysis in segmenting a corpus into groups or constellations of texts, identifying sub-corpora on internal rather than external criteria. It has made the argument that the constellations diverge from each other, both in terms of their dimensional profiles but also in terms of aspects of phraseology, illustrated here by the uses of *environment* and *climate change* in three of the constellations. Equally important, though, is the extent to which the constellations converge. This is demonstrated by the degree of similarity in the dimension profiles, and by the degree of overlap in the phraseologies discussed. We see in this successful interdisciplinary journal evidence of academic disciplines maintaining their distinctiveness but also approaching each

other, with evidence coming from corpus studies of phraseology as well as close reading of individual texts.

In terms of corpus linguistics, what this research has impressed upon us is the extent to which corpus methodologies typically prioritise divergence between corpora, whether this is used to demonstrate diachronic progression, or the difference between communities of practice, or between individual author styles. Both quantitative and qualitative methods tend to give priority to that which is maximally different. The consequence is that research findings which reveal only minor differences are interpreted as disappointing or inconsequential. In this project we try to balance the emphasis on difference with a recognition of the importance of similarity. In investigating an interdisciplinary journal we have used quite standard corpus techniques, ranging from MDA to the interpretation of concordance lines, but have offered a novel explanation of our findings in terms of both divergence and convergence between the constellations of texts.

Finally, the paper has exemplified an initial foray into research into the discourse of an interdisciplinary subject area. Whilst the possibility that interdisciplinary journals would be demonstrably distinct from monodisciplinary ones, in terms of measurable stylistic difference or a wider diversity of style, has been rejected in this instance, it has been shown that the distinct ways of approaching the topic of environmental change leave measurable linguistic traces. We have also suggested that the papers give some evidence of writers asserting both their common ground and their distinctiveness as they construct their research.

References

Biber, D. (1985). Investigating macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics*, *23*(2), 337–360. doi:10.1515/ling.1985.23.2.337

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1989). A typology of English texts. *Linguistics* 27, 3-43.

Biber, D. (1990). Methodological issues regarding corpus-based analysis of linguistic variation. *Literary and Linguistic Computing*, *5*(4), 257–269. doi:10.1093/llc/5.4.257

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam: John Benjamins.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan E. (1999). *Longman grammar of spoken and written English*. London: Longman.

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., Csomay, E., Jones, J. K., & Keck, C. (2004). A corpus linguistic investigation of vocabulary-based discourse units in university registers. In U. Connor & T. A. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 53–72). Amsterdam: Rodopi.

Biber, D., & Kurjian, J. (2007). Towards a taxonomy of web registers and text types: A multidimensional analysis. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 109–131). Amsterdam: Rodopi.

Breiman, L. E. O. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324

Charles, M. (2006). The construction of stance in reporting clauses: a cross-disciplinary study of theses. *Applied Linguistics* 27(3), 492-518.

Connor-Linton, J., & Schohamy, E. (2001). Register variation, oral proficiency sampling, and the promise of multi-dimensional analysis. In D. Biber & S. Conrad (Eds.), *Variation in English: Multi-dimensional studies* (pp. 124–137). Essex: Pearson Education.

Conrad, S., & Biber, D. (2001). Multi-dimensional methodology and the dimensions of register variation in English. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 13–42). Essex: Pearson Education.

Csomay, E. (2007). Vocabulary-based discourse units in university class sessions. In D. Biber, U. Connor, & T. A. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 213–238). Amsterdam: John Benjamins.

Fløttum, K. (ed.) (2007). *Language and Discipline Perspectives on Academic Discourse*. Newcastle: Cambridge Scholars Publishing.

Gardner, S., Biber, D. & Nesi, H. (2015). MDA perspectives on discipline and level in the BAWE corpus. Paper presented at the Corpus Linguistics 2015 Conference, Lancaster University, 21-24 July 2015. Extended abstract available at http://ucrel.lancs.ac.uk/cl2015/doc/CL2015-AbstractBook.pdf (Accessed 11 September 2015).

Gray, B. (2013). More than discipline: uncovering multi-dimensional patterns of variation in academic research articles. *Corpora* 8(2), 153-181.

Halliday, M.A.K. & Martin, J. (1993). *Writing science: literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.

Hyland, K. (2000). *Disciplinary discourses: social interactions in academic writing*. Harlow: Longman.

Hyland, K. & Bondi, M. (eds.) (2006). *Academic discourse across disciplines*. Bern: Peter Lang.

Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. In D. Biber, U. Connor, & T. A. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 73–119). Amsterdam: John Benjamins.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer.

Stirling, A. 2014. 'Disciplinary dilemma: working across research silos is harder than it looks' The Guardian online 11 June 2014. http://www.theguardian.com/science/political-science/2014/jun/11/science-policy-research-silos-interdisciplinarity Accessed 19 August 2015.