

Pre-publication version

**35 Years and 160,000 Articles: A Bibliometric Exploration of the
Evolution of Ecology**

Mark William Neff ^{a,*}, Elizabeth A. Corley ^b

^a Consortium for Science Policy and Outcomes
School of Life Sciences
Mail Code 4401
Arizona State University
Tempe, AZ 85287
mark.neff@asu.edu
Phone 480 727-0744 or 480 522 7103
Fax 480 727-8791
* Corresponding Author

^b Associate Professor
School of Public Affairs
Arizona State University
411 N. Central Avenue, Suite 450
Mail Code 3720
Phoenix, AZ 85004-0687

Will appear in *Scientometrics* Vol 8, No. 3

DOI 10.1007/s11192-008-2099-3

Pre-Publication version

**35 Years and 160,000 Articles: A Bibliometric Exploration of the
Evolution of Ecology**

Abstract

We utilize the bibliometric tool of co-word analysis to identify trends in the methods and subjects of ecology during the period 1970-2005. Few previous co-word analyses have attempted to analyze fields as large as ecology. We utilize a method of isolating concepts and methods in large datasets that undergo the most significant upward and downward trends. Our analysis identifies policy-relevant trends in the field of ecology, a discipline that helps to identify and frame many contemporary policy problems. The results provide a new foundation for exploring the relations among public policies, technological change, and the evolution of science priorities.

Keywords: ecology, bibliometrics, co-word analysis, emerging concepts, research trends

1. Introduction

Science plays a critical role in identifying and defining many contemporary environmental public policy problems. This function may satisfy some who call for increasing the role of science in the name of rational decision-making. Much of the work completed by sociologists of science from the past several decades, however, has used constructivist frameworks to describe the processes by which scientists create the information that feed into policy debates [e.g., Latour and Woolgar 1986], an approach which calls into question basic assumptions of objectivity in science. In this view, science is a social process and the truths identified by its practitioners are impossible to disaggregate from the values that contribute to their construction. Other authors have taken a soft-constructivist approach. For example, Hacking [1999] does not question the underlying truths that science reveals, but instead argues that our knowledge base would look very different if we had different research priorities. Decisions about research priorities made today will shape the very truths that science unearths for generations to come. The scientific information that helps to frame policy problems, from either viewpoint, is shaped by the societal and cultural setting in which it is produced.

This insight led Jasanoff and Wynne to argue that: “To question how *science* acquires meaning and stability, by exploring its social commitments, is to question *policy* in the same way” [1997]. Science and policy are intimately intertwined, and attempts to understand either one necessitate investigating the other. The research presented here traces research priorities in the discipline of ecology from 1970-2005 using the bibliometric tool of co-word analysis. We analyze these trends via quantitative studies of research articles published in English in ecology-specific journals. We utilize a method for identifying subjects and methods in ecology that have undergone rapid change, which allows us to identify important trends regardless of their overall magnitude. This research provides

important information that enables further exploration of the interplay between ecology and the many environmental policy problems ecology helps to identify and frame.

Authors from a variety of disciplines have explored different elements of the history of ecology, many with an eye toward the policy and cultural contexts that have led to changes in research priorities. Kwa [1987, 2005], for example, used interviews to examine how the International Biological Program (IBP) and the International Geosphere-Biosphere Program (IGBP) influenced research priorities in the field of ecology. Kohler [2002] explored how ecologists adjusted their research to maintain and build credibility in the face of pressures from both within and outside of the discipline. Other historians have taken a broader look at ecology and the factors that shape research priorities [Worster 1994; Kingsland 1995, 2005; Cooper 2003].

Though incapable of capturing the nuance of traditional historical or ethnographic methods of studying science, we believe bibliometric analysis is a powerful tool because it allows for the investigation of research priorities across an entire discipline, a perspective that eludes other forms of inquiry [Rip 1988]. Our technique allows us to track the comparative rise and fall of themes within different subfields and identify overarching trends in a way that would be prohibitively time consuming with traditional historical tools. As Callon, Law and Rip argue in the early work that laid out the theory of co-word analysis, this type of research is an attempt to “pursue the qualitative by other means” [Callon, Law, and Rip 1986].

2. Related Studies

2.1 Bibliometric Analysis

Before delving into the details of our analysis, we briefly present some background information on the methodology that we used to analyze ecological trends over time. Bibliometric analysis typically utilizes documents – primarily publications or patents – to analyze trends in science and innovation. A

variety of data analysis methods are employed during bibliometric analysis, including: co-authorship analysis, co-citation analysis (i.e., papers or authors that are often cited in tandem), and co-word analysis (i.e., words that are frequently used together in titles, abstracts, or lists of keywords) [Callon, Law, and Rip 1986; Callon, Courtial, and Laville 1991; He 1999; Leydesdorff 1997; Peters and van Raan 1993, 1993]. In this paper, we employ the tools of co-word analysis to study publication trends in the field of ecology.

2.1.1 Co-Word Analysis

Co-word analysis is based on the theory that research fields can be characterized and analyzed based on patterns of keyword usage in publications. As new concepts are proposed and new methods employed in a scientific field, authors use new words (or new combinations of words) to represent those changes. If the emerging theoretical concept or method is accepted by the broader community, those clusters of words that represent the concepts or method become ‘obligatory passage points’ for subsequent authors who wish to build upon the idea [Callon 1986]. Thus, when the scholarly literature is analyzed in aggregate through co-word analysis, ideas, concepts and methods that have been accepted by the community are identifiable due to the repeated use of the obligatory passage point terms in titles, keyword lists, and abstracts. Those ideas that are not accepted fade into the background and are not recognized in co-word analyses [Callon, Law, and Rip 1986].

Most uses of co-word analysis employ multivariate statistical techniques to identify clusters of words that appear together commonly in the literature. The source words can come from keyword lists, titles, abstracts, or other publication data fields. In past studies, researchers have used a variety of techniques to create clusters from bibliometric data, including factor analysis [White and McCain 1998], principal component analysis [Salvador and Lopez-Martinez 2000], and hierarchical clustering

[Borner, Chen, and Boyack 2003; Coulter, Monarch, and Konda 1998; Bauin et al. 1991; White and McCain 1998].

Each of these techniques organizes word groups according to the frequency with which they appear together in the database being analyzed. The resulting clusters are said to represent themes or concepts in the literature. Clusters can be created for successive time periods, allowing researchers to visualize changes in research fields over time [Ding, Chowdhury, and Foo 2001; Garfield 1994; Healey, Rothman, and Hoch 1986; Peters and van Raan 1993]. The cluster analyses conducted in this study are intended to allow to the reader to visualize themes within ecology that are dominant at different time periods – thus illuminating change over time.

Previous co-word analyses have faced a number of limitations that we attempt to transcend through the methods we employ in this paper. First, previous analyses have primarily dealt with scientific fields significantly smaller than ecology. Though smaller scale studies offer powerful windows into knowledge production, they are unable to elucidate the large-scale trends across several subfields that are of the most interest to policy makers [Law et al. 1988] and which are most likely to shape broader policy concerns outside of science; our large dataset enables us to accomplish that task. Because of our interest in large-scale, policy relevant trends in ecology, our study necessarily leaves out fine-grained trends. Secondly, whereas most previous co-word analyses have focused only on the most frequently used words in the database(s) of interest [e.g. Bhattacharya and Basu 1998; Ding, Chowdhury, and Foo 2001], our study utilizes both that approach and a technique to identify the most rapidly changing words (emerging and declining). This second analysis, similar to that introduced by Peters and van Raan [1993], allows us to isolate rapidly changing topics against the background of more stable commonly used terms. In traditional co-word analyses, these stable terms would obfuscate rapid changes in less common words. We utilize both approaches: the analysis of the most common

words to yield a snapshot of the research focus in each of the periods of our study and the analysis of rapidly changing words to allow us to identify significant trends.

Utilizing these two approaches provides an internal validity check on our results: we are able to identify the most prominent research themes at different times, track how those themes have changed in importance to the rest of ecology through strategic diagrams (explained later), and validate those trends by comparison with the emerging word analysis. An additional and powerful validity check for our results would be to compare the results of our co-word analysis to the results from a different approach to analyzing our dataset, for example co-citation analysis of highly-cited articles. A study of this sort would be a valuable future contribution to the literature.

2.2 Applications within Ecology

A number of studies have used bibliometric techniques to explore trends in a variety of ecology-related fields. Rivera [2003], for example, tracked the growth of the ‘ecosystem’ paradigm in the United States and Europe by analyzing the contents of ecology textbooks. Rodriguez and Moreira [1996] analyzed the growth of the field and its increasing complexity in Spain and five Spanish speaking Caribbean countries through an analysis of ecology dissertations. Duarte [1999] examined the growth of the field of seagrass ecology and explicitly explored the areas within the field that were expanding and changing. Carvalho and colleagues [2005] studied citation patterns to assess the impact Joseph Felsenstein has had on the field of evolutionary biology. Finally, Young and Wolf [2006] used a bibliometric review to evaluate how trends in urban ecology research align with professional society and funding agency programmatic statements of research goals. While each of these studies has contributed to the understanding of specific aspects of ecology as a scientific field, there have been few attempts to use the tools of bibliometric analysis to track changes to the field of ecology as a whole, as

we do in this paper. There are three major steps to our analysis, including: data collection, data cleaning, and data analysis (See Figure 1). These steps are described in the sections that follow.

----- Insert Figure 1 about here -----

3. Data Collection and Cleaning

The dataset for this study was constructed from the ISI Web of Science publication database. For the purposes of this study, we define the boundary of the discipline of ecology by the scholarly articles that are published within the field. To define the list of eligible journals, we consulted the Web of Science 2005 Science Edition ‘Journal Citation Reports,’ which includes a journal subject classification. Within this classification system, 112 journals were defined as belonging to the “ecology” field. Six journals listed in the classification changed names or are no longer published, and the largest number of active journals in any one year was 106 in 2005.

There are two major caveats associated with defining the field of ecology by the Web of Science journal classifications. First, this data source focuses only on published research that has been peer-reviewed; thus, the trends in the field that are captured by these data are formal, published trends that appear after a publication time lag. The time lag portion of this weakness is somewhat mitigated by the fact that we analyze only the results through the end of 2005 and data were collected in September of 2006. Second, focusing on ecology journals restricts the data to largely disciplinary articles; it does not account for articles that appear in more interdisciplinary journals, such as Science and Nature.

There are also strengths, however, to this approach for the data collection. By focusing on the peer-reviewed literature in the ecology field, we are ensuring that we capture the scholarly trends that have been vetted and accepted by the field as a whole. This is a clear strength if we are trying to capture the landscape of trends in the field of ecology, which is the focus of this paper. The field of

ecology is somewhat interdisciplinary by definition so by focusing on the journals that are classified in this field, we are not totally eliminating interdisciplinary research from our analysis. Furthermore, many of the scientific fields that have emerged from ecology during our study period, such as conservation biology and ecological economics, are represented by the journals included in our study. Lastly, several previous studies have utilized journals and/or journal classification lists to define academic fields for bibliometric analysis [e.g., Bhattacharya and Basu 1998; Schoepflin and Glanzel 2001; White and McCain 1998].

We exported the publication records for all of the English language research articles that were published in these 112 ecology journals between January 1st, 1970 and December 31st, 2005. This search, conducted in September 2006, yielded a database of 159 340 unique research articles for the 35 year time period. The bibliographic records for each article were downloaded and imported into the software package VantagePoint, version 5.0. This software was then used to verify that there were no duplicate records for articles and to generate a comprehensive list of title words. We then used the list cleaning and thesaurus features within VantagePoint to combine the plural and singular forms of words, words with alternate spellings, and hyphenated words with non-hyphenated synonyms. The first two thousand word pairings were then manually checked to validate the list cleaning methods. Any cleaning errors (such as the improper combination of ‘community’ with ‘communication’) were corrected, resulting in a final database containing 74 994 unique title words.

3.1 The Unit of Analysis

One key methodological issue in bibliometric analysis is the choice of a unit of analysis. In some past co-word studies, authors utilized article keywords or identifiers as the unit of analysis to map fields of research. There are, however, several problems with this approach. First, the means by which keywords and identifiers are selected for articles varies by journal, with some journals requiring

that authors provide the keywords and other journals assigning the task to editors or indexing services. Given this diversity of methods for article keyword selection, there has been some controversy about the legitimacy of using keywords and identifiers to map fields of research [Whittaker 1989]. A second issue is the availability of keyword data. In our dataset only a small fraction of the articles had indexed keywords or identifiers.

Words derived from article abstracts are another potential unit of analysis for bibliometric research [Peters and van Raan 1993]. The Web of Science records, however, do not reliably include abstracts prior to 1991 (James Rocco, personal communication) so we could not use these data to explore trends from 1970-2006. Therefore, given the limited set of keyword and abstract data, as well as the caveats associated with using keyword data to map fields, we decided to use article title words as our unit of analysis. Other authors have successfully analyzed research fields using title words [Bhattacharya and Basu 1998]. Whittaker [1989] has shown that title words and keywords provide similar results when used for co-word analysis, although title words tend to provide slightly less detail, a finding confirmed by Peters and van Raan [1993]. Because our study focuses on the field of ecology writ large – and we are interested in broad-scale trends – the focus on title words as a unit of analysis allowed us to analyze the appropriate breadth of data.

4. Data Analysis

In 1970, there were 918 articles in the ISI database of ecology journals; by 2005 the annual total was 10 488 articles. Therefore, when we use articles in ISI journals as a measure, the field of ecology has grown over 10-fold since 1970 in terms of the number of articles published annually. Similarly, the number of journals publishing ecology research has undergone rapid growth, increasing from 17 in 1970 to 106 in 2005. In addition, there has been a trend toward using more words per title. In our analyses we normalize to control for these trends so we can highlight changes in research

priorities rather than growth in the field (or simply increased title lengths). It is also worth noting that there has been an exponential increase in the total number of unique title words used over the study period, indicating that authors are using more diverse terms to describe their ecology research and/or are addressing more diverse topics than they formerly did.

4.1 Cluster Analysis

We divided the study period into two sub-periods in order to assess change over time [Ding, Chowdhury, and Foo 2001]. The first half of the study period (1970-1987; sub-period 1) comprises 36 844 articles and 28 842 unique words; the second half (1988-2005; sub-period 2) comprises 122 496 articles and 62 829 unique words. The most frequently used words were identified for the whole period as well as the two sub-periods using normalized word usage frequency counts. This normalization was accomplished by dividing the number of times each word was used each year by the total number of words used in all of the titles for that year. This step corrects for the substantial growth both in the number of articles published each year, as well as an observed gradual increase in the number of words per title. Thus, this normalization allows us to explore more accurately the relative importance of each term to the field of ecology each year.

4.1.1 Equivalence Matrix

The first data analysis step in co-word research is to calculate a matrix of measures of association between words in a dataset [Coulter, Monarch, and Konda 1998]. The association of words in co-word analyses is calculated based on the number of times words co-occur in individual article titles. A symmetrical co-occurrence matrix was created for each study period showing the number of times each of the 150 most frequently used words in each period were used with the other 149 words in that list. The most frequently used words were based on normalized frequency counts (as described above) but raw counts were used as input for the matrix.

An equivalence index (sometimes called a ‘strength index’ or ‘proximity index’) was calculated for each pair of words in each co-occurrence matrix [Callon, Law, and Rip 1986; Whittaker 1989; Coulter, Monarch, and Konda 1998; Cahlik 2000]. The equivalence index (E_{ij}) describes the strength of the association between words i and j in each word pair ij . It is defined as the square of the count of the number of the times two words co-occur in a title (C_{ij}^2), divided by the product of the frequencies with which each word appears in the dataset ($C_i \times C_j$).

$$E_{ij} = \frac{C_{ij}^2}{C_i \times C_j}$$

Some studies have used another index, *inclusion*, for this purpose. Inclusion is defined as the conditional probability that a document that contains word i will also contain word j [Callon, Law, and Rip 1986].

The equation for the inclusion index is:

$$I_{ij} = \frac{C_{ij}}{\min(C_i, C_j)}$$

The inclusion index is optimal for analyzing fields that are organized hierarchically, meaning that one or several high-level concepts encompass all or nearly all of the other concepts that appear at lower levels [Callon, Law, and Rip 1986].

We elected to use the equivalence index (instead of the inclusion index) because it allows for analysis of emerging and non-hierarchically organized ‘fringe concepts.’ We are interested in the entire field of ecology and its many sub-disciplines, some of which may not be organized hierarchically. Additionally, the equivalence index has the advantage of normalizing for the frequency with which each word appears in the dataset, insuring that both commonly used and less commonly used words appear in the resulting clusters [Coulter, Monarch, and Konda 1998; Callon, Law, and Rip 1986]. The equivalence index might be critiqued as overemphasizing less frequently used words, but our

'emerging words' analysis (discussed in section 4.2) allows us to explore the prevalence and dynamics of both common and less common words in greater detail.

4.1.2 Clusters

The next step in our analysis was to compute clusters. Clusters for each of the three periods (whole, sub-period 1, sub-period 2) were calculated from the equivalence matrices using Ward's agglomerative clustering algorithm [Gordon 1996; Ding, Chowdhury, and Foo 2001] with squared Euclidian distances (as recommended by Bacher 2002) using the software program SPSS. In order to allow clusters of words to emerge with minimal investigator interference, the data were not forced into a predetermined number of clusters nor were the number of words per cluster held at a constant level. Ward's method is agglomerative, meaning that each word begins as its own cluster. Words are combined to create clusters according to the frequency with which they co-occur in the dataset. This form of clustering produces a visual display of the relationships between words in the dataset, known as a dendogram. This dendogram displays the steps in the clustering process at which individual words are combined to form gradually larger clusters. Words that co-occur most frequently in the dataset are clustered early in the process and those that co-occur less frequently are clustered progressively later. Following Bacher [2002], each dendogram was inspected to identify a natural break in the steps at which words were combined into clusters. In all periods there was a natural break that resulted in a majority of clusters being between two and nine words in length. We selected clusters that were combined below this natural break for further statistical analysis.

4.1.3 Measures of Centrality and Density

Post-hoc statistical tests of centrality and density were applied to each of the clusters in order to further characterize them [Callon, Courtial, and Laville 1991; e.g., Bauin et al. 1991]. Centrality is a

measure of how often authors throughout ecology's subfields invoke the words in each cluster. It indicates how central the theme represented by a cluster is to the field of ecology as a whole. The equation that we used to calculate centrality is given below:

$$C_{Cluster} = \frac{\sum_{E_{1,1}}^{E_{ij}} E_{iw}}{[n(N - n)]}$$

where,

$C_{cluster}$ = Centrality of the cluster

E = Equivalence index of word pair link

i = First word in the pairing, internal to the cluster

w = Word in dataset, but not in the cluster

N = Total number of unique words used in titles in that period

n = Number of words in the cluster

Density, the second statistic calculated for each cluster, is a measure of the cohesiveness of the theme represented by that cluster. When an author uses one of the words in a dense cluster, s/he very frequently uses the other words in that cluster. This pattern of word usage suggests that the group of words represents a well-developed theory or method and that it has become an obligatory passage point for authors wishing to contribute to the literature. The equation that we used to calculate density is given below:

$$D_{Cluster} = \frac{\sum_{E_{1,1}}^{E_{ij}} E_{ij}}{n!/[2!(n - 2)!]}$$

where,

$D_{cluster}$ = Density of the cluster

E = Equivalence value of the link

i = First word in the pairing

j = Second word in the pairing

n = Number of words in the cluster

Pre-Publication version

Previous scholars have used a variety of equations to calculate centrality and density. For example, Callon and colleagues [1991] calculated centrality as the mean equivalence value for the six strongest external links and density as the mean equivalence value for the links within a cluster. Bauin and colleagues [1991] defined density as the sum of squares of equivalence values of internal links and centrality as the sum of squares of external link values. Our method has the advantage of correcting for the possible number of links (i.e., the number of articles and the number of words in titles in the dataset for the given period) and analyzes the full number of links rather than just the strongest.

Each cluster was mapped on a strategic diagram with axes representing centrality (x axis) and density (y axis). The axes in the strategic diagrams represent the mean values for centrality and density for the period of interest. These statistics and diagrams allow for direct comparison of the relative importance of the concepts represented by the clusters to the field of study.

In strategic diagrams, clusters with high centrality and density (quadrant 1 – upper right hand quadrant) are considered to be central to the discipline of interest and represent well-developed and cohesive concepts. Clusters with low centrality and high density (quadrant 2 – upper left hand quadrant) are said to represent peripheral but well developed concepts. Those with low centrality and density represent concepts that peripheral and undeveloped (quadrant 3 – lower left hand quadrant), and those with high centrality and low density (quadrant 4 – lower right hand quadrant) are important to the field, but comparatively undeveloped [Bauin et al. 1991; Callon, Courtial, and Laville 1991]. Thus, strategic diagrams can provide easily comparable visual representations of research priorities. We created strategic diagrams for both periods in our study to allow for visual analysis of changes in the field. Some word cluster compositions are unchanged or relatively unchanged between the two periods, indicating a certain amount of disciplinary stability around those concepts. Trends in such concepts are traceable through changes in centrality and density.

Tables 2 and 3 contain the clusters that were created for the two sub-periods, 1970-1987 and 1988-2005, respectively. The words that comprise the cluster, as well as the centrality, density, and total number of words in that cluster, are listed. Figures 8 and 9 illustrate the strategic diagrams for the two sub-periods in our study. The solid-line axes for these diagrams are the mean centrality and density for the sub-period of interest. When the means for the whole period differ from the sub-period means, the means for the full period are presented as dotted lines. Clusters that contain more than ten words and single-word clusters were left out of the analysis as outliers. The exclusion of these outliers results in a cluster label scheme which occasionally skips numbers, as may be observed in the tables and figures. The tables can be used as keys to the figures. For example, in sub-period 1 we can see that cluster number 2 is composed of three words: effects, growth, and temperature. This cluster is located in the bottom right hand corner of Figure 2 (and is represented with a '2').

----- Insert Table 1 about here -----

----- Insert Figure 2 about here -----

The strategic diagram for sub-period 1 (1970-1987; Figure 2) illustrates that during this time period there were few well-developed concepts that were central to the field. There are only three clusters located in quadrant 1, the quadrant that contains concepts that are well developed and central to the whole discipline. Listed from highest to lowest centrality, they are cluster numbers 12 (community, structure), 18 (predation, prey), and 9 (grass, prairie). Using means for the whole period (1970-2005) to define the axes, the number of clusters in quadrant 1 increases to include clusters 15 (selection, habitat), 8 (distribution, abundance) and 2 (effects, growth, temperature). The clusters with the highest densities (i.e., those representing well developed concepts) tended to have low centrality during sub-period 1, meaning that they are unimportant to the field of ecology on the whole. They also contain fairly general terms. Examples include clusters 27 (factors, affecting) and 36 (conditions,

under). Cluster 32, containing the words ‘small’ and ‘mammals’ is the single densest cluster in this sub-period, meaning that when an author uses one of these terms s/he very frequently uses the other. The centrality for this cluster, however, is near the low for that sub-period suggesting that it is comparatively unimportant field of ecology on the whole.

----- Insert Table 2 about here -----

----- Insert Figure 3 about here -----

The strategic diagram for sub-period 2 (1988-2005; Figure 3) reveals significant changes from the first. Perhaps the most immediately obvious change is that there are 13 clusters in quadrant 1 for this time period, compared with three from the previous sub-period. The number increases by two when the axes are defined by the whole-period means. The single most central cluster in sub-period 2 comprises the words ‘forest’ and ‘tropical.’ The density of this cluster is the second highest for the period, indicating that the tropical forest research theme is extremely well developed and important to many authors in the field of ecology.

4.2 Emerging and Declining Concepts

The clustering method described above is a powerful tool for identifying dominant research themes and characterizing change over time. However, this technique only allows for the analysis of a limited number of words, reducing its usefulness for identifying research themes that are undergoing rapid change. Concepts that are not adequately described by the 150 most commonly used words in any period are excluded from the above analysis. Other authors have similarly limited previous co-word analyses to the most commonly used words, and their analyses suffer from the same shortcoming [e.g., Bhattacharya and Basu 1998; Ding, Chowdhury, and Foo 2001].

We believe that identifying emerging themes within a rapidly growing and changing field like ecology may provide valuable insights into how and why disciplines evolve. For that reason, we apply

a method similar to that used by Peters and van Raan [1993] to identify and analyze words that have undergone the most significant changes in usage patterns. These words are then clustered to illustrate visually the concepts they represent. To identify emerging and declining words, annual word usage counts were normalized by dividing the number of appearances of each word each year by the total number of title words for the year. This normalization, which was also used to identify the most frequently used words for the above described clustering, corrects for the rapidly increasing number of articles published as well as an observed trend toward using longer titles in our study period.

The normalized values for each word were summed for each of the two sub-periods of our study, 1970-1987 and 1988-2005, and the difference between those sums was then calculated. We ranked the values of those differences to identify the words that have undergone rapid upward and downward trends¹. The 150 words that increased the most and the 150 words that declined the most were then subjected to cluster analyses as described above. Emerging words were clustered according to their co-occurrence patterns in the second sub-period because the concepts represented by those took shape in that period. By that same logic, declining words were clustered according to their co-occurrence in sub-period 1. Emerging word clusters are shown in Table 3; declining word clusters are shown in Table 4.

----- Insert Table 3 about here -----

----- Insert Table 4 about here -----

The resultant clusters were subjected to analysis of centrality and density and then mapped on strategic diagrams. The emerging word strategic diagram, Figure 4, reveals that two general themes

¹ Our technique privileges words that are used most frequently to the possible exclusion of words undergoing more rapid change but which are used comparatively infrequently. Had we chosen to do so, this bias could have been eliminated by dividing the calculated difference between the normalized word usage in period 1 and 2 by the total normalized frequency count for that word overall. Alternatively, we could have used a mean comparison t-test (following Peters and van Raan, 1993) or used the Pearson correlation coefficient between year and word usage counts to identify words that have undergone statistically significant changes. We chose not to normalize in those manners, however, because we are most interested in large-scale changes to the field of ecology rather than rapid but comparatively smaller scale trends.

have become increasingly important to the field of ecology. First, there are three different clusters with above average centrality containing words related to genetic research: number 8 (genetic, variation, structure); number 34 (DNA, mitochondrial); and number 21 (microsatellite, characterization, loci). The fact that three separate clusters relating to a group of techniques each have above average centrality indicates that genetics and related tools have increased in importance to the field of ecology over the study period and are now heavily used by the authors of ecology journal articles. Similarly, there are two separate clusters with above average centrality with words related to climate change: cluster 11 (changes, climate); and cluster 42 (CO₂, elevated). The cluster containing ‘forest,’ ‘tropical,’ and ‘rain’ (number 2) is the most central cluster among those constructed from the emerging word list.

----- Insert Figure 4 about here -----

To explore these trends further, we subjected the individual words comprising these clusters to further analysis. The trends of the individual words that make up clusters number 34 and 21 are shown in Figure 5. From this finer analysis it becomes clear that the words that comprise cluster 34 are quite new to the field of ecology, with the word ‘DNA’ gaining popularity in the late 1980s and ‘mitochondrial’ rising in importance beginning in the mid 1990s. The trends are similar but more pronounced for the words in cluster 21: ‘microsatellite,’ ‘characterization,’ and ‘loci’ all experienced a rapid increase in usage beginning in the mid 1990s. The word ‘microsatellite,’ a genetics term associated with heredity and population studies, was unknown to the lexicon of ecology until that time.

----- Insert Figure 5 about here -----

The trends of words from the two clusters related to climate change are shown together in Figure 6. There is noticeable annual variability in the usage of each of these terms, but there is a marked increase in frequency of each beginning in the late 1980s and early 1990s.

----- Insert Figure 6 about here -----

Clusters were calculated from the list of words showing the most dramatic downward usage trends as well. These clusters are displayed in Table 4 (mentioned earlier) and Figure 7.

----- Insert Figure 7 about here -----

There are markedly fewer clusters in this sub-period because many of the words clustered together in a single cluster of 115 words (excluded from the analysis as an outlier). This large cluster emerged because there are no strong co-occurrence patterns of the words therein. The only cluster in the quadrant indicating strong concept cohesion and importance to the field of ecology as a whole (quadrant 1) is number 9 (deer, whitetailed, mule). Two related clusters (number 4: production, primary; and number 13: photosynthesis, net) have above average centrality. Articles related to each of these themes became progressively less common over the study period.

5. Limitations of the Approach

In our analysis, co-word techniques allowed us to gain traction in the problem of understanding trends in the large and highly dynamic discipline of ecology a phenomenon with significant policy relevance. As Callon, Law and Rip [1986] suggest, it provides a way of doing qualitative research by other means and open up policy relevant research questions that are not amenable to other approaches. That said, however, the literature on co-word analysis has identified a number of limitations to the approach. We are able to mitigate several of those limitations, such as the limited number of words that can be analyzed at any given time, through deliberate selection of methods. Our analysis complimented a standard co-word analysis of the most commonly used words with an analysis of words undergoing rapid change. In this section, we discuss another significant challenge to the co-word approach and our response to those critiques.

Several authors have suggested that co-word analysis is weakened by the plurality of meanings of any given word at a single point in time and is vulnerable to changes in the meaning of words over

time. In a 1997 study, Leydesdorff utilized co-word analysis to study the ways in which words were used in individual papers and concluded that language is too dynamic for co-word analysis to be an effective way of mapping changes in science. While the changing nature of language over time and the variety in ways in which individual words are used at any given time are significant concerns for co-word analysis generally, we feel that these problems are most formidable when the techniques are applied to finer-scaled analysis of smaller disciplines or sub-disciplines. The large-scale trends that we are able to identify, such as the rapidly growing importance of genetic techniques across ecology's subdisciplines, are robust to the multiple meanings of a given genetics-related term and slight changes in the meanings and uses of those terms over time. We are not aware of other techniques that would enable analysis of trends across a discipline of this size.

Some observations in this study clearly capture changing scientific diction, but changes in language in our field are interesting in and of themselves because of their implications for environmental and other policies. Biodiversity is a concept that came into common usage during our study period. 'Biodiversity' was never discovered; rather, it is a word that was invented to encapsulate a variety of other concepts and ethical norms [Takacs 1996]. Its rapid rise to prominence within ecology might therefore be best described as a linguistic change, and even though its meaning clearly has changed over the past several decades, the ascension of 'biodiversity' as a concept within ecology marks one of the most significant trends in that field and one that has been mirrored by significant attention in the broader policy world.

We agree with Leydesdorff's proposition that the variety of ways in which words are used in different contexts complicates attempts at linguistically-based artificial intelligence, but we do not feel it seriously limits the utility of co-word analysis to identify broad-scale changes in a research field over time in the aggregate. In this paper, we are not interested in subtle changes in the definitions of words;

rather, we are interested in changes in the themes of research because that is what is most relevant to policy research [Law et al. 1988], and because of the course-grain of our investigation we believe it to be less susceptible to subtle changes in word usage than are smaller scale studies interested in fine-grain trends.

A related phenomenon that can potentially hinder attempts to use title words to characterize scientific disciplines is the fact that authors can use title words (and indeed, any words in a text) strategically to tie the article to a hot topic, thereby attempting to make the article more likely to be published, read, and cited. Although this ‘re-labeling’ can be seen as disguising the true content of the article, such a move would simply serve to reify and reinforce the importance of the concept (the obligatory passage point) that author was using to bolster his or her chances of being published and read. The obligatory passage point is thereby made more obligatory for future authors, thereby hastening actual change in research priorities.

6. Discussion and Conclusions

To conclude our paper, we return to our original research question: “How have research priorities within the field of ecology changed over the past three decades?” Historian Donald Worster remarked that in the postwar years:

Ecology achieved intellectual sophistication, academic prominence, and financial security . . . but also lost much of its coherence. It broke down into a cacophony of subfields, including ecosystematists, populationists, biospherians, theoretical modelers, forest and range managers, agroecologists, toxicologists, limnologists, and biogeographers [1994].

While the context of this quote suggests that his point is that ecology lacks a unitary policy voice, an argument we do not address, he makes a broader claim about the structure and evolution of ecology,

which can be treated as a hypothesis that we have tested here. Our analysis supports the first part of his claim: Ecology has grown immensely since 1970, possibly because of the increasing prominence and financial security he describes. In our analysis that growth is most visible in terms of the number of ecology journals and the articles they publish. The evidence on the second part of his claim is mixed. On one hand, several of the themes that now dominate ecology cut across the ‘cacophony of subfields’ he describes, suggesting some growing methodological unity in the field. On the other hand, there has been a proliferation of ecology journals over the past 35 years, each of which specializes in different aspects of ecology. We explore the dominant trends that emerge from our analysis, their correspondence with Worster’s claims, and their importance below.

The field of ecology has matured and developed significantly since the 1970s and early 1980s. For example, the strategic diagram for the first period (1970-1987) illustrated that there were only three central and well-developed clusters within the research literature in the field of ecology before 1988. This lack of well-developed and central research foci indicates that the discipline was comparatively undifferentiated during that period and that few themes had emerged as important across ecology. The strategic diagram in the second period (1988-2005) showed a significant increase in the number of central and well-developed clusters in the field. Whereas there were three such clusters in the first time period, there were thirteen well-developed and central clusters within the field of ecology during the second time period of our analysis. While the conclusions that we can draw about these changes are somewhat general, the results do indicate a proliferation of comparatively well-developed concepts over time.

Some of the clusters from the second study period appear to more or less correspond to the subfields Worster describes. For example, period 2 cluster 3 (population, genetic, dynamics) and 14 (distribution, abundance) can be associated with populationists. Among other examples, period 2

cluster 6 (model, predicting) can be linked to his subfield ‘theoretical modelers.’ These results seem to support Worster’s suggestion that ecology has split into a cacophony of subfields. The high centrality of these clusters suggests, however, that these concepts and methods are used by authors across all of ecology. These central clusters are evidence of some level of conceptual and methodological unity in ecology.

Additionally, the emerging word analysis that we utilize shows that the maturation of ecology has included an increasing focus on subjects such as climate change and genetics that cut across established ecology subfields. Some of these emerging crosscutting trends are enabled by new technologies. For example, genetics-related terms in clusters that are related to new technologies (cluster 21: microsatellite, characterization, loci; cluster 34: DNA, mitochondrial) have undergone much more rapid change than those that are less technology-dependent (cluster 8: genetic, variation, structure). Based on the fact that there are two clusters with high centrality that are closely linked with genetic technologies, we can assert that researchers across ecology’s subfields utilize these technologies. Though these findings do not contradict Worster’s claim that ecology has splintered into ‘a cacophony of subfields,’ they do suggest that these subfields share interest in subjects such as climate change and the technology-oriented methods of genetics.

These trends, while interesting on their own, are critical background information for future efforts to understand the factors that shape research priorities. As argued in the introduction, science plays a significant role in identifying and framing environmental problems. Because the research that is conducted influences the types of problems science identifies, knowledge of how research priorities emerge is critically important to any understanding of the role that science plays in society. Sociologists and anthropologists of science have explored the factors that shape individual researchers’ choices of research problems [Zuckerman 1978; Ziman 1987; Gieryn 1978], but there have been

Pre-Publication version

inadequate efforts to explore the factors that shape entire disciplines [Kleinman 1991]. The work presented here, with its approach to identifying emerging trends in ecology, provides necessary background information for further exploration of questions such as: 1) What role do funding agencies play in generating scholarly interest in new subjects and methods? 2) How do technological change, research priorities, and policy changes interact? 3) How do political trends (e.g., interest in tropical forests and endangered species), research funding, and research relate to each other over time? 4) Are there common traits to the trends that emerge? For example do they begin in certain journals? Or, are they nurtured by particular funders or universities? Our analysis and the database we have constructed allows us to further explore these questions. Research like that presented in this paper, which could be conducted on any discipline with readily available publication records, opens up numerous other studies to better understand research processes and the policies that impact research priorities.

Acknowledgements

This material is based upon the work supported by the National Science Foundation under Grant No. 0345604. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We are indebted to Search Technology, Inc for access to the software program Vantage Point (www.thevantagepoint.com). We gratefully acknowledge comments on previous drafts of this paper by Bethany Cutts, Kate Darby, Kris Gade, Maya Kapoor, Ann Kinzig, Libby Larson, Thad Miller, Charles Perrings, Dan Sarewitz, and Shade Shutters. Shivshankar Venugopal offered critical support in the calculation of centrality and density.

7. References

- Bacher, J. 2002. Cluster Analysis. Nuremberg: University of Erlangen-Nuremberg.
- Bauin, S., B. Michelet, M. G. Schweighoffer, and P. Vermeulin. 1991. Using bibliometrics in strategic analysis: "understanding chemical reactions" at the CNRS. *Scientometrics* 22 (1):113.
- Bhattacharya, S., and P. K. Basu. 1998. Mapping a research area at the micro level using co-word analysis. *Scientometrics* 43 (3):359–72.
- Borner, Katy, Chaomei Chen, and Kevin W. Boyack. 2003. Visualizing Knowledge Domains. *Annual Review of Information Science and Technology (ARIST)* v37 p179-255 2003.
- Cahlik, Tomas. 2000. Comparison of the Maps of Science. *Scientometrics* 49 (3):373.
- Callon, M. 1986. The Sociology of an Actor-Network: The Case of the Electric Vehicle. In *Mapping the Dynamics of Science and Technology*, edited by M. Callon, J. Law and A. Rip. London: Macmillan.
- Callon, M., J. Law, and A. Rip. 1986. How to Study the Force of Science. In *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*, edited by M. Callon, J. Law and A. Rip. London: Macmillan.
- . 1986. Qualitative Scientometrics. In *Mapping the Dynamics of Science and Technology. Sociology of Science in the Real World*, edited by M. Callon, J. Law and A. Rip. London: Macmillan.
- , eds. 1986. *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. London: Macmillan.
- Callon, M. N., J. P. N. Courtial, and F. N. Lavoie. 1991. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics* 22 (1):155-205.
- Carvalho, Priscilla, Jos Diniz-Filho, Felizola Alexandre, and Luis Mauricio Bini. 2005. The impact of Felsenstein's "Phylogenies and the comparative method" on evolutionary biology. *Scientometrics* 62 (1):53.
- Cooper, G. J. 2003. *The Science of the Struggle for Existence: On the Foundations of Ecology*. Cambridge: Cambridge University Press.
- Coulter, N., I. Monarch, and S. Konda. 1998. Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science* 49 (13):1206-23.
- Ding, Y., G. G. Chowdhury, and S. Foo. 2001. Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management* 37 (6):817-42.
- Duarte, C. M. 1999. Seagrass ecology at the turn of the millennium: challenges for the new century. *Aquatic Botany* 65 (1):7-20.
- Garfield, E. 1994. Scientography: Mapping the Tracks of Science. *Current Contents* (November).
- Gieryn, Thomas F. 1978. Problem Retention and Problem Change in Science. *Sociological Inquiry* 48 (3/4):96.
- Gordon, A. D. 1996. Hierarchical classification. In *Clustering and Classification*, edited by P. Arabie, L. J. Hubert and G. D. Soete. River Edge, NJ: World Scientific Publishing.
- Hacking, Ian. 1999. Weapons Research. In *The Social Construction of What?:* Harvard.
- He, Q. 1999. Knowledge Discovery through Co-Word Analysis. *Library Trends* 48:133-59.
- Healey, Peter, Harry Rothman, and Paul K. Hoch. 1986. An experiment in science mapping for research planning. *Research Policy* 15 (5):233.
- Jasanoff, S., and B. Wynne. 1997. Science and Decision Making. In *Human choice and climate change. Volume 1: The societal framework*, edited by S. Rayner and E. L. Malone. Columbus, OH: Battelle.

- Kingsland, Sharon E. 1995. *Modeling Nature: Episodes in the History of Population Ecology*: University of Chicago Press.
- . 2005. *The Evolution of American Ecology, 1890-2000*. Baltimore: The Johns Hopkins University Press.
- Kleinman, D. L. 1991. Conceptualizing the Politics of Science: A Response to Cambrosio, Limoges and Pronovost. *Social Studies of Science* 21 (4):769-74.
- Kohler, Robert E. 2002. *Landscapes and Labscapes: Exploring the Lab-Field Border in Biology*. Chicago: University of Chicago Press.
- Kwa, Chunglin. 1987. Representations of Nature Mediating Between Ecology and Science Policy: The Case of the International Biological Programme. *Social Studies of Science* 17 (3):413-42.
- . 2005. Local Ecologies and Global Science: Discourses and Strategies of the International Geosphere-Biosphere Programme. *Social Studies of Science* 35 (6):923-50.
- Latour, B., and S. Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*: Princeton University Press.
- Law, J., S. Bauin, J. P. Courtial, and J. Whittaker. 1988. Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics* 14 (3):251-64.
- Leydesdorff, Loet. 1997. Why words and co-words cannot map the development of the sciences. *Journal of the American Society for Information Science* 48 (5):418-27.
- Peters, H. P. F., and A. F. J. van Raan. 1993. Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Research Policy* 22 (1):23.
- . 1993. Co-word-based science maps of chemical engineering. Part II: Representations by combined clustering and multidimensional scaling. *Research Policy* 22 (1):47.
- Rip, Arie. 1988. Mapping of science: possibilities and limitations. In *Handbook of Quantitative Studies of Science and Technology*, edited by A. F. J. van Raan. Amsterdam: Elsevier.
- Rivera, Adolfo Cordero. 2003 Trends in the evolution of ecology: "Spain is different". *Web Ecology* 4:14-21.
- Rodriguez, Ketty, and J.A. Moreira. 1996. The growth and development of research in the field of ecology. *Scientometrics* 35 (1):59-70.
- Salvador, M. R., and R. E. Lopez-Martinez. 2000. Cognitive structure of research: scientometric mapping in sintered materials. *Research Evaluation* 9 (3):189-200.
- Schoepflin, Urs, and Wolfgang Glanzel. 2001. Two decades of "Scientometrics". An interdisciplinary field represented by its leading journal. *Scientometrics* 50 (2):301.
- Takacs, D. 1996. *The idea of biodiversity: philosophies of paradise*: Baltimore: Johns Hopkins University Press.
- White, Howard D., and Katherine W. McCain. 1998. Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science* 49 (4):327-55.
- Whittaker, J. 1989. Creativity and Conformity in Science: Titles, Keywords and Co-Word Analysis. *Social Studies of Science* 19 (3):473-96.
- Whittaker, J., J. P. Courtial, and J. Law. 1989. Creativity and Conformity in Science: Titles, Keywords and Co-Word Analysis. *Social Studies of Science* 19 (3):473-96.
- Worster, D. 1994. *Nature's Economy: A History of Ecological Ideas*. Second edition ed. Cambridge: Cambridge University Press.
- Young, R. F., and S. A. Wolf. 2006. Goal attainment in urban ecology research: A bibliometric review 1975–2004. *Urban Ecosystems* 9 (3):179-93.
- Ziman, J. M. 1987. The problem of "problem choice". *Minerva* 25 (1):92-106.

Pre-Publication version

Zuckerman, Harriet. 1978. Theory Choice and Problem Choice in Science. *Sociological Inquiry* 48 (3/4):65.

Figure 1. Method flow chart

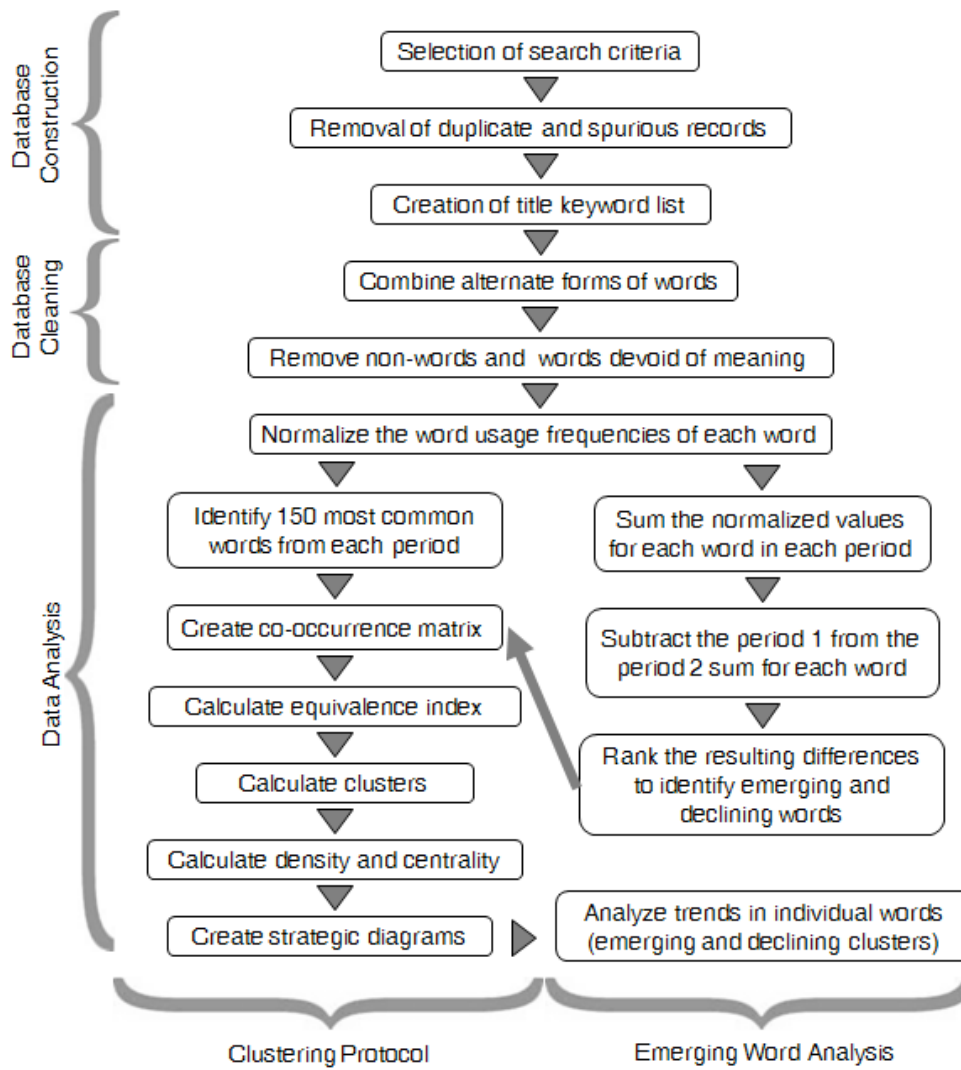


Figure 2. Period 1 (1970-1987) Strategic Diagram

(Note: Solid lines depict means for this period; dashed lines are means for the whole period (1970-2005))

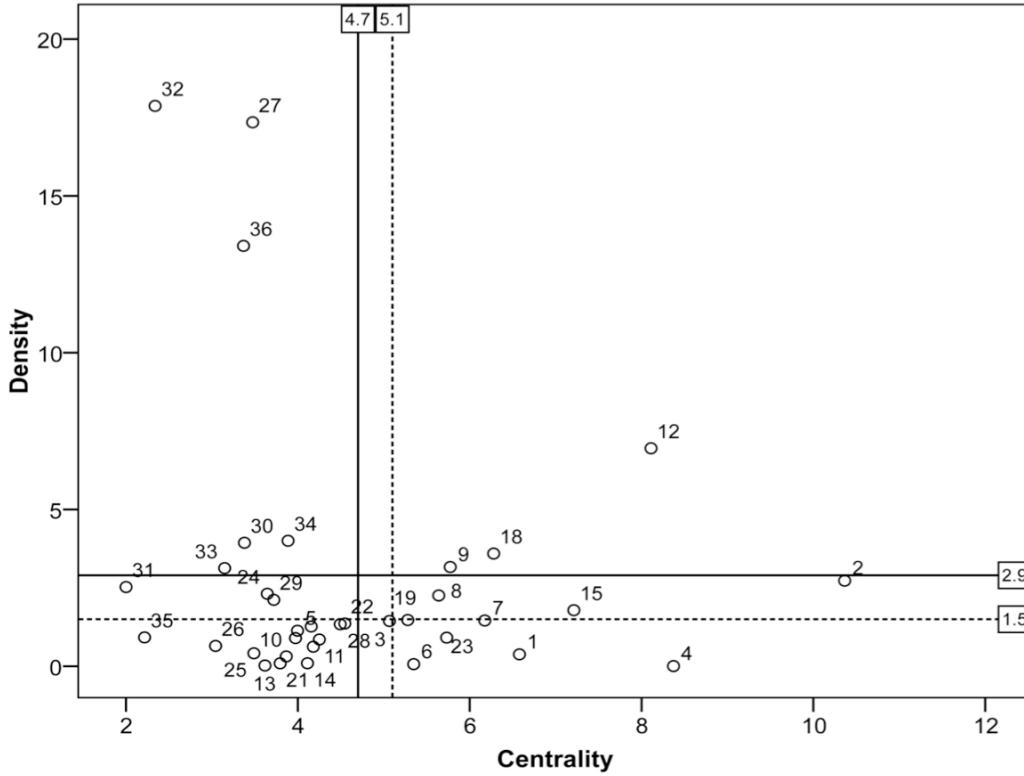


Figure 3. Period 2 (1988-2005) Strategic Diagram

(Solid lines depict means for this period; dashed lines are means for the whole period (1970-2005))

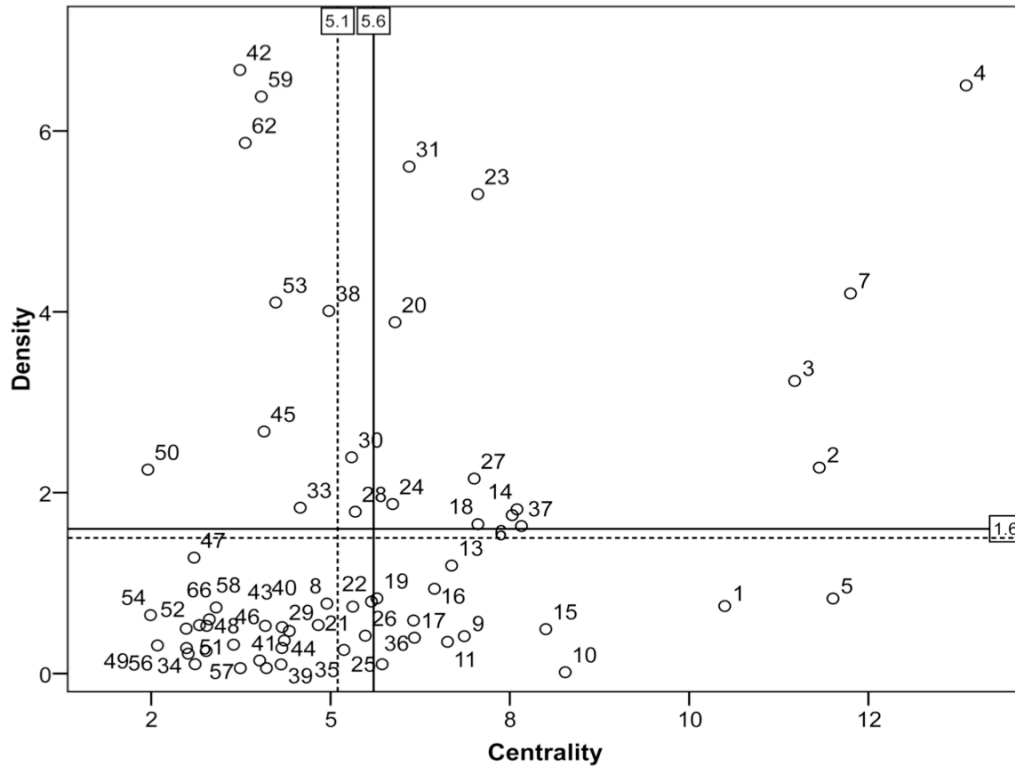


Figure 4. Emerging Words Strategic Diagram

(Solid lines depict means for emerging word clusters; dashed lines are means for period 2 (1988-2005))

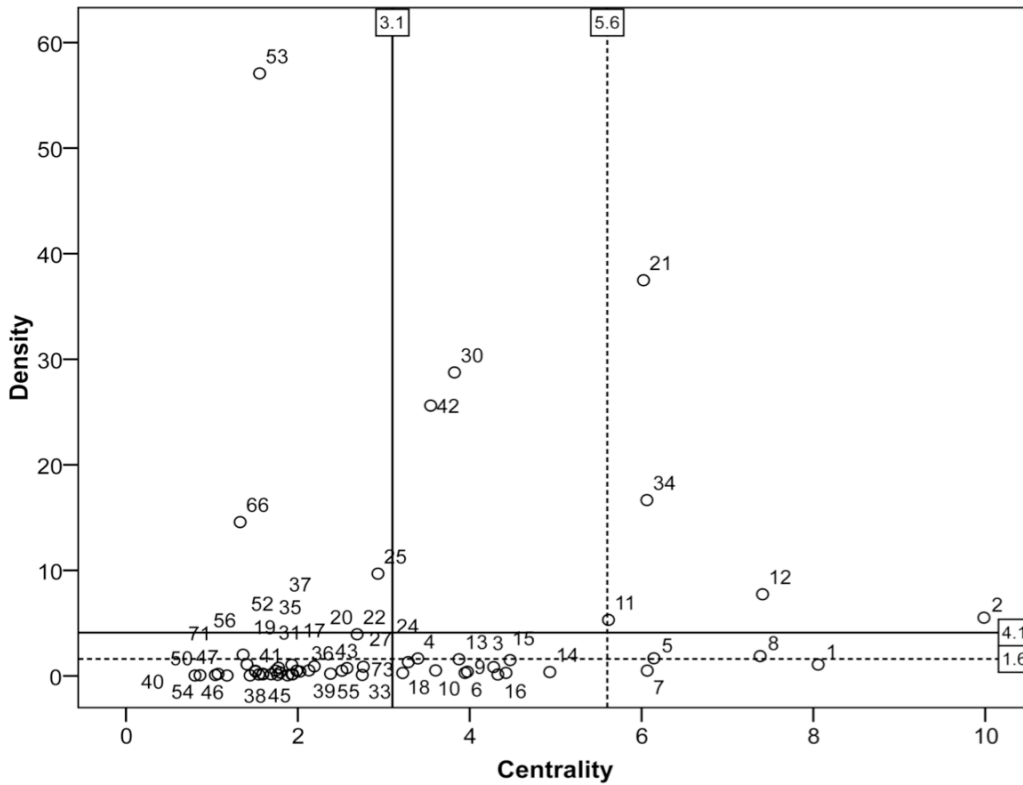


Figure 5. Normalized Word Use Frequencies for Emerging Words: Clusters 21 (microsatellite, characterization, loci) and 34 (DNA, mitochondrial)

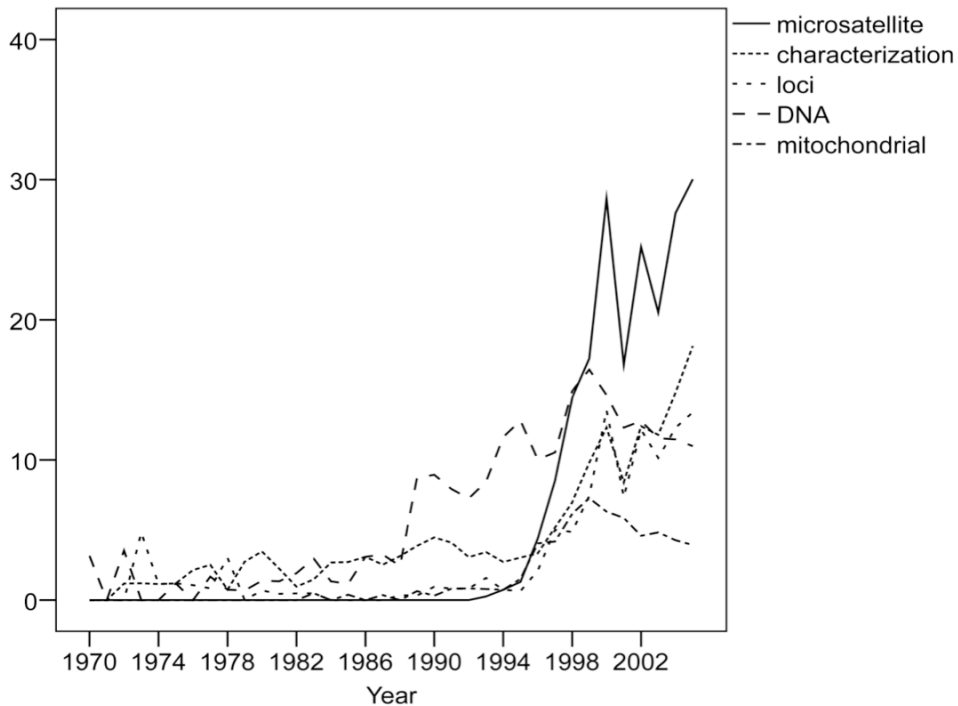


Figure 6. Normalized Word Use Frequencies for Emerging Words: Clusters 42 (CO2, elevated) and 11 (climate, changes)

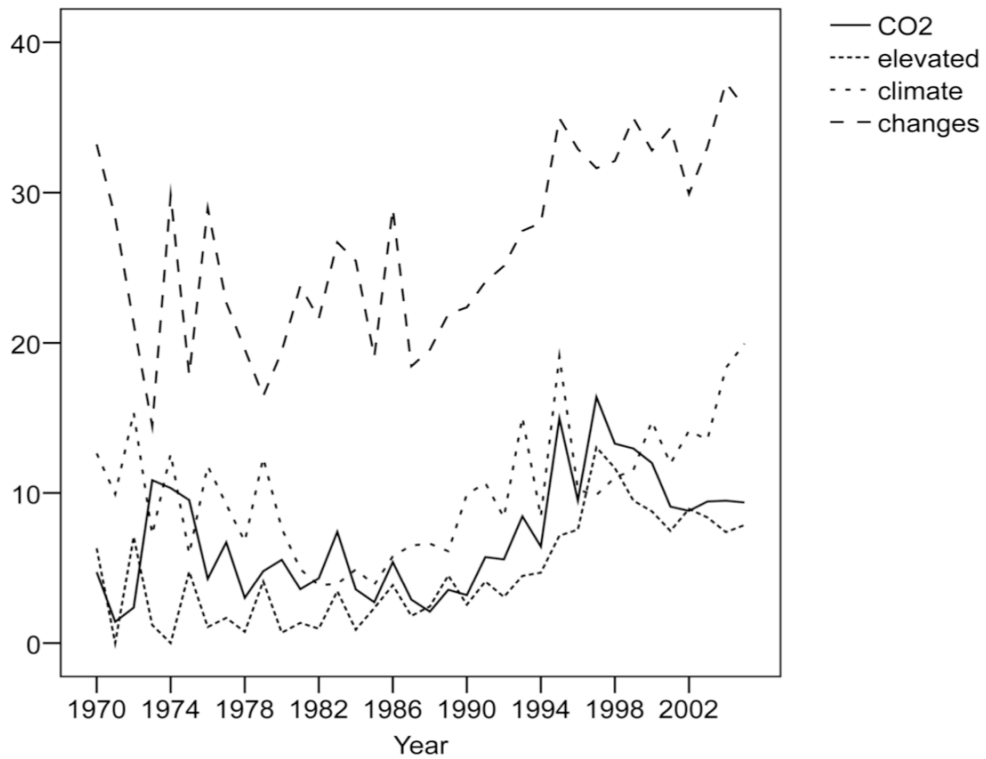


Figure 7. Declining Words Strategic Diagram

(Solid lines depict means for declining word clusters; dashed lines are means for period 1 (1970-1987))

