

BOOK APPENDIX IV: Does the Choice of Textual Analysis Software Affect the Results?

I. Introduction

Quantitative empirical researchers are long accustomed to receiving suggestions and advice on how to improve their results—for example, by gathering more data, introducing more variables or refining the existing ones, or perhaps changing the model specification. For researchers analysing textual data, some of these same suggestions might hold, but more often one is asked about the quality of textual analysis software itself. Unlike statistical analysis software—much of which is fairly standard in its output, given the strong foundations in statistical methods—textual analysis software can be more specific to the task and/or the researcher, thereby raising the question: to what extent are the empirical findings a product of the software, rather than the data? There are at least three problems underlying this query. First, the statistical and theoretical foundations for textual analysis do not adhere to a single framework, and are thus open to dispute. Second, software packages often fall into one of two categories—proprietary or open-source. The processing methods for the former are invariably opaque while they are usually transparent for the latter. Social science researchers understandably argue that all algorithms, assumptions and processes of text analysis software should be fully transparent (Lowe 2003)—which implies that they are freely available. There is clearly a tension here between market forces and the development of scientific knowledge (hardly unique to textual analysis), which leads to a third problem: the growing plethora of incompatible textual analysis software which produce fundamentally different types of results. With such obstacles to achieving robust, defensible results from textual analysis, what is the way forward?

One answer to achieving a reasonable threshold of robustness is to ask, do my data look different when I examine them from different perspectives or using different methodological toolkits? If so, one may well have less confidence in the initial approach. If not—if the same fundamental results emerge again and again—the researcher can be fairly certain that she is on solid footing. Looking at data from different perspectives is an increasingly attractive way forward for a number of social scientists (Klüver 2009; Lahlou 2011; Lowe and Benoit 2012), and is the approach taken here.

This paper has two parallel aims—one methodological and one more substantive. With respect to methodology, the basic motivation of this paper is to assess the extent to which different automated content analysis software yield broadly similar results, when applied to the same corpora. The two corpora analysed here originate from our book that seeks better to understand deliberations on US monetary policy (Schonhardt-Bailey 2013). In our book, we sought to capture the content and substance of the discourse in the House and Senate hearings on the Federal Reserve's Monetary Policy Report over a 33 year period (1976-2008). We

examined whether committee members understood, (a) the underlying objectives of monetary policy, (b) the ways in which the Fed could be judged to be meeting these objectives, and (c) the implications of the Fed's conduct of monetary policy for other critical issues areas—like the labour market and other aspects of the wider economy and policy. We also explored whether senators and congressmen differed in their discourse in the monetary policy oversight hearings.

Our empirical analysis of the textual data in *Deliberating American Monetary Policy* employed Alceste as the automated content analysis software. There are clear advantages to this software, as we explain in our book appendix I (see <http://mitpress.mit.edu/DAMP>). There are also disadvantages—chief of which is that the software is proprietary (although a limited open-source version of it now exists¹). The software was conceived initially in a public research institute (CNRS) and is widely used in the policy and academic communities. The practical implementation of the method remains somewhat opaque, although a recent work makes explicit its linguistic and methodological assumptions and offers alternative algorithmic operationalizations (Chartier and Meunier 2011).

This appendix presents a more specific challenge to the software by employing two other automated content analysis software on the transcripts of the congressional hearings—T-Lab and Dtm-Vic. The former is proprietary software (similar to Alceste) while the latter is open-source. These two software packages were chosen because they both employ methods and functions that approximate some of those found in Alceste and so provide enough common functionality to allow comparisons of the results, without in any way serving as actual replicas of Alceste. In Section III, we outline the basic approaches of each software package (more complete descriptions of each may be found easily elsewhere²).

Apart from the comparison of results from the three software packages, this paper has a second goal that derives from the first—namely, to uncover new findings and/or new insights pertaining to the deliberations of members of Congress as they conduct oversight. In *Deliberating Monetary Policy* we observed a number of key characteristics on the oversight behaviour of legislators in the House and Senate banking committees. Here the task not only is to challenge the initial findings but also to push beyond them: in short, what more can we learn from adding new methodological perspectives?

In the next section, key findings from *Deliberating American Monetary Policy* (abbreviated below as *DAMP*) are summarized; Section III provides a brief description of Alceste, T-Lab and Dtm-Vic; Section IV presents the findings from the different methodological perspectives; and Section V concludes, particularly in addressing the question of whether this added effort is a worthwhile and reasonable approach for researchers of textual data.

II. Key Findings from Alceste: Congressional Deliberation on Monetary Policy

[Figures 1 & 2, about here]

In *DAMP* the overwhelming finding was that legislators in both the House and Senate banking committees largely failed to meet a minimal standard of deliberative discourse (with “minimal” defined loosely as an *exchange* of ideas and arguments which focus on the same subject matter (in this case, monetary policy)). Such an exchange was largely absent from the oversight committee hearings, as the Fed chairman and legislators tended to talk past one another rather than to one another. With the exception of a few key individuals (mostly committee chairmen and ranking members), legislators simply did not engage with the Fed chairman on the particulars (the guts) of monetary policy, and consequently, his remarks on topics such as monetary aggregates and the US real economy stood relatively unchallenged. As illustrated in dendrograms of the classes (Figures 1 and 2), we saw a clear cleavage between the areas of focus of the Fed chairman and members of either the House or Senate banking committee. In this sense, there was no real difference between the House and Senate committees. We did, however, find that senators were more likely to question the Fed chairman on institutional issues (e.g., appointments to the Fed and intergovernmental relations between Congress, the administration and the Fed) and on issues relating to foreign economic policy, like US competitiveness in the world economy. In the aggregate, moreover, fiscal policy received more attention in the House hearings relative to the Senate hearings (with the exceptional year of 2005, when spending on the Iraq war captured the attention of the Senate committee). While these differences of substance are noteworthy, the key point here is that with respect to monetary policy per se, committee members of both chambers were about equally willing to leave the details to the Fed chairman. This raises questions regarding the quality, and even relevance, of congressional oversight on monetary policy—as we discuss in our book conclusion.

We did find, however, that committee members—and particularly Democratic members—tended to focus on the implication of monetary policy for labour markets. In this respect, committee members were able to hone in on one aspect of monetary policy for which they could score points with their constituents: jobs and employment. Broadly speaking, committee members tended to speak to particular themes outside the realm of monetary policy for which they could score points with their constituents—e.g., institutional challenges to the independence of the Fed (in the era of high inflation) or acquiring advice/recommendations from Greenspan on politically sensitive topics like Social Security reform, education, income inequality or energy (in the era of low inflation).

[Figure 3 – about here]

Importantly, the backdrop for these hearings is a period of significant change in US monetary policy from the mid 1970s to 2008 (as seen in the inflation rate in Figure 3), and corresponding to these changes is a pro-cyclical political response by Congress. That is, congressional oversight is stringent and sceptical when the Fed is perceived as delivering a failing monetary policy; and it is lenient, passive (and even effusive) when the Fed’s policy actions appear successful. While this cyclical nature may seem reasonable at first glance, one might also question whether legislative oversight should strive to eliminate cyclical

tendencies (particularly in the wake of a financial crisis that was preceded with a period of laxer oversight).

Our period of analysis begins with the high inflation of the 1970s and the Fed's ineffectual policy response, continues to the "revolutionary" actions of the Volcker Fed which began to rein in the inflationary spiral, followed by a long period (from the later 1980s into the new century) of stable low inflation alongside a growing economy, and ending with the early days of the financial crisis. We found that in bad economic times of high inflation, the tendency is to score points by shifting blame to the Fed, while in good economic times, legislators seek not only to praise the Fed chairman but also to seek his support for specific (non-monetary) policy stances. Moreover, over this same period, central banks throughout the world came to accept and endorse the notion that a policy of low inflation—together with central bank independence—is a fundamental precursor to stable economic growth (known as the "low inflation consensus"). In the US, Paul Volcker's campaign to persuade members of Congress of the importance of sustaining a policy of low inflation appears to have created a window of opportunity for the Fed to acquire a reputation for committing credibly to delivering on low inflation, thereby ushering in a period in the 1990s and early 2000s when inflation became essentially a non-issue in oversight hearings. The topic of inflation resurfaced in 2006 not because inflation had suddenly escalated, but rather because Ben Bernanke—who was well-known for his advocacy of inflation-targeting—became Fed chairman.

These findings appeared consistent with the actual first-hand experiences of key monetary policymakers and staff from the Fed, as well as legislators and congressional staff from the two congressional committees (as gleaned from our in-depth interviews and reported in chapter 5 of our book). The question here is, does the use of *different* textual analysis software alter the basic story of (1) members of Congress who did not engage with the Fed chairman on monetary policy per se, (2) marginal, but not fundamental, differences between the thematic focus of senators versus congressmen in their respective banking committees, and (3) a change over time in the apparent understanding of the contribution of sustained low inflation to US macroeconomic policy (especially a growing acceptance and lack of challenge to the low inflation consensus)?

III. Automated Content Analysis, from Three Perspectives

Much could be said regarding the proliferation of text-mining software and its application to political texts, but such a wide-ranging overview is not the intention here. Rather, the three software used in this paper all fall into a sub-set—i.e., automated (or computer-assisted) content analysis. Within this category are at least two distinct types of software: topic models and thematic-based approaches (elsewhere the relative merits of these two different types are examined (Schonhardt-Bailey, Yager et al. 2012)). All three software under investigation here (Alceste, T-Lab and Dtm-Vic) are thematic-based approaches to automatic content analysis and all use mixed-methods (including word co-occurrence

analysis, correspondence analysis, and so on). Hence, the assumption is that speakers convey meaning in a thematic fashion, and so it is not just the words that help to classify content, but also the *context* in which the words appear.

a. Alceste

The first software—Alceste—considers the text as a large matrix of co-occurrences between lexical forms, and processes it with multivariate techniques. A key feature of Alceste is that it can be used to identify the speakers' tendency to articulate particular ideas and arguments—ideas and arguments which can then be correlated with characteristics of the speaker (e.g., in political texts—the name of speaker, party affiliation, constituency characteristics and so on).

In the first book appendix of *DAMP* we provide a more detailed description of the algorithms and their rationale, but in brief, Alceste operates in four steps: it parses the vocabulary (step A); it transforms the corpus into a sequence of Elementary Context Units (ECUs) containing words (or more exactly lemmas) and operates a descending classification which produce stable classes of these ECUs, leaving what does not fit in these classes “unclassified” (step B); it operates a series of statistical characterizations of the classes (typical words, typical sentences, crossing variables, providing χ^2 values, etc.) (step C), which enable the analyst to operate interpretation (step D). The interpretation consists in attributing meaning to the “lexical world” that is latent in each class based on these statistical results. The software thus follows an iterative process where the descending hierarchical classification method decomposes the classes until a predetermined number of iterations fails to result in further divisions.³ The result is a hierarchy of classes, which may be schematized as a dendrogram (or tree diagram). Correspondence analysis may also be used to examine relationships between classes, as well as between classes and characteristics of speakers.

The software provides a number of tools for the researcher to interpret each class, and two tools are particularly useful—the characteristic words and the characteristic phrases.⁴ Both are ranked in order of χ^2 significance, to allow a clearer understanding of the terms and phrases which predominate in each class.

b. T-Lab

Whereas Alceste uses descending hierarchical classification, T-Lab employs an ascending hierarchical clustering approach, in which the bisecting K-means algorithm allows one to repeat the bisecting of clusters until the desired number of clusters is reached (Steinbach, Karypis et al. 2000; Savaresi and Boley 2001). Simply put, this means that Alceste begins with the total set of context units in a corpus, which constitutes the first class. Using a recursive algorithm, the program then attempts to partition that class into two further classes that are each as homogenous as possible and as different from one another. As noted above, the descending hierarchical classification method (following an iterative process) decomposes the classes until a predetermined number of iterations fails to result in further significant divisions. In contrast, T-Lab's ascending approach essentially works from bottom up, using co-occurrence more flexibly—as directed by the user. The package allows a

representation of the corpus through a minimum and maximum number of clusters (3 and 50, respectively). Hence, the user is able to fix the number of cluster partitions (although the clustering algorithm stops if further partitions fail to meet the statistical criteria).

The practical implication of this approach is that the operational relevance of thematic categories (“classes” [Alceste]; “clusters” [T-Lab]) is more discretionary in T-Lab, meaning that once the maximum number of partitions is identified for a particular corpus, the user may then conduct further analysis at any point in the hierarchy of clusters. For example, if the corpus comprises ten clusters, one might choose to conduct analysis on all ten clusters or perhaps at a more aggregate level of three clusters, depending on what seems more relevant to the researcher.

Details of the software are easily available elsewhere (both in the user’s manual, but also in the bibliographic list available on the T-Lab website: <http://www.tlab.it/en/bibliography.php>), and so the discussion here will focus on some characteristics that underpin the uniqueness of this approach. As the architect of the software discusses at length (Lancia 2007), T-Lab is anchored in *abductive* reasoning, which allows one to abduce a single explanation—or category of analysis—from an abundance of contextual units. The basic idea is that meaning and thematic similarities may be achieved from recognizing patterns in word co-occurrence. Using Lancia’s examples, one might observe the following sentences:

That student is reading the software manual.
Waitress! Can you bring us the menu?
The film was greeted with applause by the press.

Using co-occurrence analysis, he extracts just the content words and re-writes these:

manual, reading, software, student
bring, menu, waitress
applause, film, greeted, press

Clustering algorithms for co-occurrence analysis then allow one to *abduce* a shared meaning, because words that co-occur “in similar contexts tend to have similar meaning” and “documents that contain similar word patterns tend to have similar topics” (Lancia 2007: 25). Importantly, the dictionary of content words is critical to this process. If for example, the first sentence about the student was followed by a second one—“She sat in the jungles of South America”—and “jungle(s)” and/or “South America” were not included as content words, the interpretation of the combined context unit would acquire a significantly different meaning. This helps to show a key difference between Alceste and T-Lab, which matters for the interpretation of the results: whereas T-Lab facilitates and prompts the researcher to modify the dictionary of content words and lemmas during the process of analysis, Alceste relies more on an automatic dictionary of content and auxiliary words, although the user can refine the lemmatization process by careful pre-coding of the corpus. A good example of this in the

corpora for this paper is the word “Fed”, which is short for Federal Reserve. Alceste’s lemmatization process interprets “Fed” as the past tense of “feed” and so this must be corrected to avoid mis-classification (that is, the lemmatization process must be supervised).

A second key difference is that the two software deal with the issue of where to “cut” the context unit quite differently. If, for example, we again use the example about the student along with the added sentence, we must decide where to cut the text into context units. Should we cut the context unit at the end of the first sentence or combine the two sentences into a single unit? The two software approach this issue quite differently. For Alceste, the process of descending hierarchical classification is conducted twice, using different context unit lengths (e.g., once with 12 words and again with 20 words). Only those context units that are successfully classified in both processes are retained for analysis. In T-Lab, the user must decide the length of the context unit (sentence, “chunks”, or paragraphs). So once again, the onus is on the researcher to direct the analysis.

c. Dtm-Vic⁵

Dtm-Vic (**D**ata and **t**ext **m**ining - **V**isualization, **i**nference, **c**lassification) is open-source software (<http://www.dtmvic.com/>), developed by Ludovic Lebart. Broadly speaking, Dtm-Vic provides a number of the same tools available in Alceste and T-Lab—e.g., hierarchical clustering, partitioning, and correspondence analysis. Also similar to the other two software, it allows textual data (speeches, documents) to be agglomerated (or tagged) by categorical variables (year, role of the speaker [such as committee chair or member], party affiliation, and so on), which in turns allows one to produce and analyse lexical tables (contingency tables, or “categories x words”).

All three software use lemmatization rather than stemming, in order to reduce forms to a common base. (For instance, *predictable*, *predicted*, *predicting*, *predicts* is reduced to *predict*; and *go*, *going*, *gone*, *went* is reduced to *go*). Whereas stemming usually implies a process of chopping off ends of words, lemmatization is generally more refined in that it involves the use of a vocabulary and an analysis of words in order to reduce them to their dictionary form (known as the *lemma*). All three software convert lemmas into types (articles, prepositions, pronouns and other function words). But whereas Alceste and T-Lab have in-built lemmatization processes, for Dtm-Vic this is achieved by a plug-in software (in the case of this paper, the free and widely used software TreeTagger [<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html>] and for the Windows interface: <http://www.smo.uhi.ac.uk/~oduibhin/oideasra/interfaces/winttinterface.htm>]).

As author of many books in statistics and textual analysis (e.g., (Lebart, Morineau et al. 1984; Lebart and Salem 1994; Lebart, Salem et al. 1998), Lebart is particularly sensitive to validation techniques (Lebart 2003). Thus, a unique feature of Dtm-Vic is the availability of *Bootstrap* methods as a way to validate the obtained results. The *Bootstrap* techniques and their adaptations to textual data allow us to decide whether the observed patterns are significant in a statistical sense, as opposed to being the results of some random noise. They

produce confidence areas (ellipses or convex hulls) around points plotted on the principal axes maps whether those points represent words or texts. These confidence areas thus discourage the over-interpretation of the graphical patterns.

The *bootstrap* technique gives rise to an unresolved issue in textual analysis software—namely, what ought to be the appropriate measure for “goodness of fit” to the data (somewhat akin to the coefficient of determination in regression analysis or in other analyses, Pearson’s chi-squared or likelihood ratio tests)? This is relevant since we want to know how well our resulting classification or clustering results reflect the underlying structure of the textual data. For Alceste, the double classification procedure produces a percentage, which denotes those ECUs that have been successfully classified (twice) in the same class (as a robustness check). This percentage of ECUs successfully classified is an approximate for a goodness of fit measure. For T-Lab, the classification rate for the ECUs is linked to the a priori specification of the length of the ECU (sentences, chunks, paragraphs), and then reflects the full set of clusters available for analysis. (For instance, in the House corpus analyzed below, the specification of “chunks of sentences” yields a classification rate of 95%, while specifying “sentences” yields a slightly lower one of 93.6%.)

The problem is that Alceste and T-Lab are both “unsupervised”, meaning that the classes or clusters are built from the data. (In “supervised” analyses, classes or categories are pre-existing. Classification is then a “diagnostic”, for which the researcher may evaluate the model’s error (mis-classification rate) as well as its success (classification rate).) From a statistical perspective, the computation of a classification rate in unsupervised analyses is not possible (given that it is endogenous to the data—i.e., there is no independent reference from which to compute success and failure). The problem deepens when one considers that statisticians do not even agree on the definition of a cluster itself (is it defined by areas of low density around the borders, or by the existence of a dense nucleus?)—but this extends beyond the remit of this paper. The point is that there is no agreed upon method for clustering, and so the approach taken by Dtm-Vic is to assess the stability of patterns observed in the textual data with the use of bootstrap techniques, rather than cluster validations. These techniques will be illustrated below.

IV. Congressional Oversight Hearings on Monetary Policy: From Three Different Perspectives

a. Comparison of Themes across Software

We return now to our two corpora—congressional hearings on monetary policy in the House and Senate, over the period of 1976-2008. To what extent do the results from T-Lab and Dtm-Vic confirm our findings that (1) members of Congress did not engage with the Fed chairman on monetary policy per se, (2) marginal, but not fundamental, differences existed between the thematic focus of senators versus congressmen in their respective banking committees, and (3) there was a change over time in the apparent understanding of the contribution of sustained low inflation to US macroeconomic policy (especially a growing

acceptance and lack of challenge to the low inflation consensus)? In order to assess our three questions, we must address two fundamental issues. First, we must establish that the overall clustering of thematic classes is consistent with our earlier findings, and subsequent to that, that the key variables (Fed chair, member, committee chair) show the same basic misalignment in thematic classes. For this, we rely on T-Lab (and for the moment set aside the lack of consensus among statisticians regarding the definition of clusters). A second issue is whether the hearings' discourse convey the same (or similar) pattern of ideational change over time—and for this, we use Dtm-Vic.

[Tables 1 through 4, about here]

Tables 1 through 3 present the basic statistics for both the House and Senate hearings, as produced by Alceste, T-Lab and Dtm-Vic. Rows two through four in each table give the total word count (which varies slightly, given differences in the structure of the data for each software) and then details on the words analysed and frequency thresholds. Row five is particularly important, as it denotes the pre-existing variables or categories that each software analyses. For Alceste, this appears to be a large number (237 and 129), but in fact is just four variables—year, role of committee member, party affiliation and name. The inflated tally reflects the large number of different names for the committee members, which Alceste analyses as unique variables. For T-Lab, the number of categories for any given variable is limited to 150, and so the name variable is dropped from the analysis. And, for Dtm-Vic, this analysis uses just the year variable (though other variables could have been analyzed). Reassuringly, the number of speeches remains constant among all three software (row six).

In Tables 1 and 2, we observe the so-called classification rate for both Alceste and T-Lab, and notwithstanding our comments above concerning the ambiguity of this measure, both software appear to achieve high rates (91%-92% for Alceste, and 95% for T-Lab, though as noted above, this reduces slightly if the context units are analysed as sentences rather than chunks of sentences). In row nine of the first two tables, we observe the number of classes or clusters. For Alceste, the descending classification algorithm obtains nine stable classes. T-Lab's ascending classification obtains a maximum of ten clusters, allowing the user to specify the number desired for further analysis. For comparability with Alceste, the analysis is partitioned at nine clusters. The final row then gives the relative weights of each of the classes/clusters, using the characteristic words and ECUs to provide the labels. In Table 3 (Dtm-Vic), this final row corresponds to each category of the year variable, thus summing to twenty for each set of hearings. Labels for each year are derived from characteristic words only. We will return to discuss these below.

Table 4 condenses Table 1 and 2 into a more manageable format, allowing a direct comparison of the themes found in both Alceste and T-Lab. Columns two and three give the results for the House hearings, and columns three and four give results for the Senate. There are six themes that are common to both sets of results: (1) fiscal policy, (2) regulation, (3) monetary aggregates, (4) real economy, (5) uncertainty and challenges to monetary policy, and (6) the policy process—which includes the institutional relationship between the Federal Reserve and Congress (e.g., oversight) as well as the macroeconomic “mix” between fiscal

and monetary policy. Bold font indicates a direct match in the substance of the themes, while a class given in italics signifies that the content is very closely related, but has marginal differences. So, for the first three themes—fiscal policy, regulation and monetary aggregates—we find essentially the same classes/clusters in both sets of results for both the House and the Senate.

For the real economy theme, there is broad commonality in the House results. At first glance, the same appears to be the case for the Senate, although in T-Lab this emerges as two clusters—“Real Economy and the Labour Market” and “Change in Variables Relating to the Real Economy”. This latter cluster illustrates a functional difference between Alceste and T-Lab, which deserves further investigation. Using the example of the real economy theme, there is a direct parallel in the “US Real Economy” class in Alceste, and the “Change in Variables relating to the Real Economy” cluster in T-Lab, as both contain the top key words *percent, rate, quarter, year, and average*. However, the classification of the “Labour Market” discourse is more of a challenge—perhaps given its links to other areas of the economy, including the real economy, training and education of the workforce and overall competitiveness of American workers. For Alceste, we obtain a separate class that focuses on “Education, Training and US Competitiveness in the Labour Market”, with the following as the most representative lemmas: *job, education, skills, people, class, inequality, wealth, trade, Americans, China, and college*. The top ECU (key words in bold) is from Senator Brown, in February 2007:

... the uncertainties that **middle class families** face are not the uncertainties that the columnist that Senator Bennett **mentioned** and **others** and economists **worry** about as often perhaps as they should. I know and appreciate your acknowledging the widening **gap** of income in our **society**. I **commend** you for adding your voice to that discussion. I agree with you that we should **look** at ways to improve **education** and **training** of our **citizens**, but I do not think that is nearly enough. **Globalization** has had a **tremendous impact** on **workers** in this **country**, on **communities**, on teachers, on firefighters, on cities’ ability to deliver services to their constituents. There is no question that good paying **manufacturing jobs** have gone offshore. Fourteen years ago, the **trade** deficit in this **country** was 38 billion dollars. Today, announced just this week, it exceeds 760 billion dollars. George Bush the first said that a 1 billion dollar **trade** deficit translates **into** 13, 000 **lost jobs**. You do the math. Of course, we must **trade** with the **world**. The question is not if we will **trade** with other countries; rather, it is how we will **trade** with them and who will **benefit**.

For T-Lab, the labour market discourse is subsumed in a larger discourse about the domestic real economy, with the following top key lemmas: *people, know, talk, program, work, happen, think, and job*. The top ECU is from Senator Sarbanes in July 1992:

There is really no **problem**. It is like the **president** who says, the economy’s getting **better** and the **American people** do not **know** it. The fact of the matter is that the **American people know exactly** what is **happening** and they **know** there is a lot of economic **trouble** out there.

Thus, perhaps owing to the differences in the lengths of elementary context units, T-Lab clusters tend to reflect word co-occurrence where the words are situated in closer proximity, whereas the longer ECU in Alceste encompasses more distant linkages between jobs and other factors, like globalization, trade competitiveness and so on. This class appears in the bottom of Table 4, where the classes and clusters that have no direct parallel across both

Alceste and T-Lab are listed. In these bottom rows, it is apparent that Alceste has classified themes that contain inherent linkages across ideas or concepts. For instance “Members of Congress Prompting Fed Chair on Non-Monetary Policy Issues” relates to topics like Social Security, energy policy and so on; “Capital Inflows, Exchange Rate, Current Account Deficit” links these economic aspects of the economy; “World Economy and the US External Balance (i.e., Trade and Current Accounts)” again embodies related economic concepts; and finally, as explained above, “Education, Training and US Competitiveness” is closely linked to the labour market. The two clusters that are unique to T-Lab exhibit concepts that are immediately more closely linked: both “Bank Lending and Credit Creation” and “Inflation and Prices” are clusters containing ideas/concepts that are directly related (very simply, bank lending creates credit, and inflation increases prices).

In sum, Table 4 illustrates that core themes in the discourse on monetary policy emerge, regardless of which software is used. However, where the form of argumentation is ideationally more complex—meaning that ideas and arguments are bridged across more distantly related topics, like funding for education and the labour market—Alceste appears to capture this as a thematic class, while T-Lab does not. In substance, we can say that six core themes—fiscal policy, regulation, monetary aggregates, real economy, uncertainty/challenges, and the institutional/policy process—are reasonably robust. Measuring statistically how legislators and the Fed chairman talk about jobs is more tricky, given that jobs may be linked (rightly or wrongly) to a wide variety of possible growth agendas. Hence, one might say that the labour market theme is more susceptible to variations in methodological approaches.

We can also, at this point, address at least one of our three key questions—namely, whether differences between the discourse in the House and Senate hearings are consistent across software. In short, we want to confirm whether the House members were somewhat more likely to focus on fiscal policy, and whether senators were more concerned with the world economy and US competitiveness, as well as with Fed-Congress relations (institutional issues). On fiscal policy, both software give a marginally greater weight in the House than the Senate, so there is consistency in this finding. On the foreign economic policy theme, T-Lab does not appear to confirm senators’ greater interest in this topic—though this could relate to the discussion above concerning Alceste’s ability to capture more distantly related ideas. Lastly, the finding on Fed/Congress relations is less clear. Whereas the result in Alceste was that congressmen talked more about policy process (including the question and answer format) and senators more about governance issues (appointments and inter-institutional issues), T-Lab appears to capture more of a focus on committees (congressional and the Fed’s monetary policy committee—FOMC) by congressmen and a conceptual focus on the mix between fiscal and monetary policy by senators. It is probably safe to conclude that both software have some difficulty in capturing with absolute clarity the discussions among legislators and the Fed chair regarding the process of congressional oversight and the institutional relations between the Fed and Congress. One reason might be that both oversight and inter-institutional relations ebbed and flowed throughout the three decade period, and both were subject not only to uncertainty but also to political machinations. So, perhaps it

should come as no surprise that automated textual analysis software finds it challenging to pin down these topics with a great deal of certainty.

b. Variables and Themes (or, Did Legislators and the Fed Chair Talk Past One Another?)

[Figures 4 and 5, about here]

We turn now to examine whether the role of the committee member—as congressional committee chair or member; or as Fed chairman—relates to the area of discourse. In particular, we want to know whether politicians and the Fed chairman tended to talk past one another, as suggested by our earlier results. Dendrograms of the Alceste classes (Figures 1 and 2) illustrate a conspicuous cleavage in the discourse between politicians and the Fed chairman. Figures 4 and 5 present similar dendrograms for the T-Lab clusters of the House and Senate hearings. All ten available clusters are mapped, but in each case, it is easy to see that partitioning the clusters into nine merely aggregates the two clusters that are most closely linked in terms of word co-occurrence. The clusters for the House hearings divide reasonably well into themes that relate to monetary policy and those relating to other areas of policy or process. The thematic mapping for the Senate hearings is not as tidy, with monetary policy topics more widely spread throughout the dendrogram.

[Figures 6, 7 and 8, about here]

A better way to assess the variables is with bar charts, as given in Figures 6, 7 and 8. Using the nine partition model in T-Lab, Figures 6 and 7 pinpoint the top three clusters, by committee chairman, Federal Reserve chairman, and committee member. In Figure 6, *inflation and prices* is the primary theme for both the committee chairman and Fed chairman. Second to that for the Fed chairman is *uncertainty*, followed by *monetary policy*. This finding squares very well with our results from Alceste—i.e., the Fed chairman uses the hearings to discuss inflation, uncertainty, and of course, monetary policy itself. Beyond the shared focus on inflation, the committee chairman is more interested in talking about *monetary aggregates* than about uncertainty. The next two (equally weighted) topics for the committee chairman pertain to *committees* (congressional and FOMC) and the process of oversight, as well as *fiscal policy*. In contrast, committee members focus on fiscal policy first and foremost. After that, are the two themes on uncertainty and committees. The clearest cleavage is between the Fed chairman and committee members, with the former focused on inflation and monetary policy and the latter on fiscal policy and the oversight process of committees. Committee chairmen are a mix between the Fed chairman and committee members—i.e., they offer exchange to the Fed chairman on the substance of monetary policy, but also retain a focus on the more political dimensions of fiscal policy. This corresponds to our summary findings—i.e., the majority of the committee members go to the hearings to talk about something other than monetary policy, with the exception of a few key members, like the committee chairmen. In the Senate hearings (Figure 7), there is again a difference between the areas of focus for the Fed chairman and committee members: the Fed chairman discusses monetary aggregates and inflation, while committee members focus on labour market aspects of the

real economy and on fiscal policy. And again, the committee chair bridges the divide by discussing both monetary aggregates and labour market aspects of the real economy.

Together, Figures 6 and 7 tell both an inter- and intra-institutional story in monetary policy oversight hearings: (a) central bankers tend to talk about the guts of monetary policy while the committee members use the hearings to discuss other matters, e.g., the labour market and fiscal policy; and (b) within the oversight hearings themselves, the committee chairman is the bridge between the Fed's focus on monetary policy and committee members' concern with politically sensitive (and more transparent) issues like jobs, government spending, taxes, and so on.

What about partisan differences among members? Do these affect the discourse? In *DAMP* we found marginal partisan differences in the discourse—e.g., Democrats (in both the House and Senate) were more inclined to focus on jobs. For simplicity, we present T-Lab results of partisanship for the Senate only. In Figure 8, two observations are readily apparent. First, the cleavage in themes between the Fed chairman (designated as NoP, or No Party) and committee members of both parties is striking. *Both* Democrats and Republicans are chiefly concerned with labour market issues in the real economy, while the Fed chairman discusses monetary aggregates and inflation. So, with respect to labour market aspects of the real economy, partisanship does not appear to be relevant in the Senate. However, in the House hearings, the discourse of the vociferous Independent Bernie Sanders focuses predominantly on the *labour market* – and is entirely consistent with our findings and discussion of this unusual committee in *DAMP*. Second, Republicans are, however, far more predisposed to discussing the structure of banking and bank regulation (and de-regulation) than are Democrats, which is not something that was evident in Alceste.

Where does this leave us, then, in responding to our first two questions, concerning the lack of engagement of legislators with the Fed chairman on the details of monetary policy, and the marginal differences between the discourse in the hearings across chambers? Both software are consistent in finding that the Fed chairman and committee members diverge in their areas of focus—leading one to conclude that in oversight hearings, central bankers and politicians tend to “talk past one another”. Both software also find that the committee chair is unusual in his engagement with the Fed chairman on the details of monetary policy, though in T-Lab, the results show more clearly that committee chairs appear to balance talking the language of monetary policy with talking the language of politics.

As for the marginal differences across chambers, the two software packages find a somewhat greater emphasis on fiscal policy in the House. However, in Senate, Alceste finds a greater focus both on inter-institutional issues and the world economy. We attribute this to methodological variations in how each software assesses complex arguments (given differences in the lengths of context units).

In short, replicating the analysis of the aggregate House and Senate hearings does not substantially change our initial story, although it does raise the question as to how best to measure (statistically) complex arguments that bridge disparate concepts and idea.

Clearly we have only scratched the surface, both in our analyses of the corpora and software. In *DAMP* we explore deliberation from a broader array of perspectives—including comparisons between how politicians discuss monetary policy in congressional oversight hearings and how central bankers discuss monetary policy; and by using different methodological approaches such as elite interviews, regression analysis and closer visualization of the data with correspondence analysis, box plots, histograms and so on. But, duplicating such extensive analysis in T-Lab is well beyond the scope of this paper—and indeed would try the patience of even the most devoted student of monetary policy deliberations.

c. Monetary Policy Discourse over Time

[Figure 9, about here]

Turning now to our results from Dtm-Vic, we examine changes in the discourse on monetary policy oversight over three decades—from 1976 to 2008. In *DAMP*, we explored these changes in depth, but here key findings are summarized. From Figure 3, we know that US monetary policy underwent a sea change from the 1970s to the late 1990s: whereas in the earlier period of high inflation, the Fed was seen as failing in its management of monetary policy, by the late 1990s, its apparent success was evidenced by a growing economy amidst persistent low inflation. The backdrop for this policy shift is a paradigm shift in how central bankers throughout the world interpreted the role of inflation in the economy. Reining in inflation became the primary objective for monetary policy, and independent central banks were the key institutional device for delivering this policy. In the US, there was likewise a change in terms of the how best to control inflationary tendencies. Managing inflationary tendencies through seeking to control the money supply (using monetary aggregates as a measure) gave way to seeking to influence the inflation *expectations* of consumers, investors, businesses, and so on.

In the discourse on monetary policy in congressional hearings, we observe a number of changes in response to this evolving backdrop. On the conceptual understanding of inflation itself, we see in the congressional hearings a shift away from discussions of monetary aggregates (with inflation more directly the product of the money supply) and towards the real economy (inflation as shaped by the balance of supply and demand in the economy and the expectations of agents). We illustrate this in *DAMP*, using bar charts of the monetary aggregates theme over time (Figure 9). After the mid-1980s, discourse on this theme disappears. In Table 3, we see the same result in Dtm-Vic (we have indicated this theme in red font). Moreover, in *DAMP* we found that legislators appeared to have accepted the low inflation consensus as the objective of monetary policy over time, as evidenced (in part) by the finding that inflation itself became a non-issue in the oversight hearings—i.e., disgruntlement over inflation had disappeared. Of course, this corresponds with an apparently successful monetary policy during the Greenspan years of the 1990s and early 2000s. So, as inflation appeared to be well under control, the discourse shifted to other areas (jobs, Social Security). The exception to this finding comes at the end of our period of study when Ben Bernanke sought to focus the oversight hearings on to the low inflation objective of monetary

policy, though with little sign that members of congress embraced it with enthusiasm as a subject of discussion. This pattern of discourse—with inflation largely absent from the discourse in the 1990s and early 2000s—is consistent with the key topics by year, as given in Table 3 (with “inflation” in blue font). These findings accord with our broader interpretation of monetary policy oversight as being pro-cyclical, that is, less attentive/critical in the good times and exceedingly so in the bad times.

[Figures 10 through 15]

Turning to the correspondence analysis of the hearings, and using the single variable of the year of the hearing, we see a clear pattern over time in Figures 10 through 15. In the House hearings (Figures 10 through 12), we plot first the results from an analysis that cross tabulates all the years and the 1242 lemmas that appear at least 50 times (Figure 10, and reported in Table 3). There is a clear progression across the graph, from right to left, with 2008 standing conspicuously apart from 2006 and 2007.

In Figure 11, we increase the frequency threshold to include only the 303 lemmas that appear at least 300 times. We see roughly the same pattern across the graph, starting with the earlier years on the right and moving to the later years on the left. Using this same high frequency threshold, we replicate the correspondence graph, adding thematic labels to the six conspicuous year clusters (Figure 12). In the later 1970s, discourse focused on the problem of high inflation—understood and measured as a product of the supply of money (monetary aggregates). The Volcker Revolution of 1979 marks a shift in the Fed’s management of inflation (as we explain in *DAMP*), thereby creating a new cluster in the early- mid-1980s. By the early 1990s (Cluster 3), inflation is well under control, and committee members in the House draw Greenspan into discussions of fiscal policy. The period of low inflation continues—even in the midst of a growing economy—which leads to a focus on the US (real) economy in the late 1990s (Cluster 4), then returning in the early years of the new millennium to fiscal policy and jobs (Cluster 5). So far, we have a clear story in the discourse that accords well with (a) the pro-cyclical nature of congressional oversight (i.e., to discuss inflation in monetary policy in the bad times and then areas of more conspicuous policy outcomes in the good times—like fiscal policy); and, (b) a particular tendency of Greenspan to focus on productivity in the US economy in the later 1990s (a topic we discuss in *DAMP*). The final sixth cluster is uniquely situated apart from the remaining five clusters—which distinguishes its unusual discourse. Starting in 2006, we see a fundamental shift away from themes focusing on the US economy, fiscal policy, and the labour market. These are replaced with issues of central importance to the financial crisis—namely, housing, the crisis of credit and the management of risk. Notably, the year in which the crisis hit severely—2008, with the collapse of Bear Sterns, Lehmans, AIG and so on—is set apart from 2006 and 2007, as one might expect given the cataclysmic events of that year.

Correspondence analysis for the Senate oversight hearings appears in Figures 13 through 15. In the first two graphs (which depict the results reported in Table 3), the contingency table cross-tabulates the twenty years and the 441 lemmas that appear at least 200 times in the Senate corpus. Once again, we observe a right-to-left movement in the discourse, by year. In

Figure 14, four distinct year clusters are identified. The first cluster resembles the content of the first two clusters in the House hearings, with the focus on monetary aggregates, controlling inflation and the Volcker Revolution. In the mid-1980s, however, the Senate discourse differs considerably from that in the House. Cluster 2 (1984-85) falls at the time of an overvalued dollar (a product of the Fed's higher interest rates under Paul Volcker) which contributed to import competition for US industries. Congress thus faced lobbying for protectionist legislation, and the Reagan Administration sought to fend off this pressure by negotiating a depreciation of the dollar relative to key currencies (resulting in the Plaza Accord in 1985). In contrast to House representatives, senators in the Banking Committee engaged in discussions with Volcker about issues relating to the dollar's devaluation (including reducing US current account deficit). This result supports our findings in *DAMP* that the Senate committee tended to focus relatively more on foreign economic policy than did its House counterpart. Cluster 3 in the early 1990s—concerned with the labour market and fiscal policy—parallels the third cluster in the House. Finally, Cluster 4 illustrates a broader shift in discourse over time towards the US economy and productivity, as well as non-monetary issues such as funding Social Security.

Finally, Figure 15 represents the selection of seven years and eleven key words, in order to illustrate the confidence ellipses derived from *Bootstrap* resampling. These ellipses are drawn both for year points and word points. The ellipses relative to year points are very small, owing to the sizes of the texts corresponding to each year. They define confidence areas constituting a validation of the pattern of years. The confidence areas relative to words show that their locations on the plane are far from being random. The ellipses relative to the words *economy* and *investment* are however overlapping; hence one cannot reject the hypothesis that their chronological distributions are different. Broadly speaking, we see a thematic shift from a focus on *inflation and interest rates* to one focusing more on *jobs, US economy, Social Security* and ultimate, *financial regulation*.

V. Conclusion: Is This a Sensible Way Forward?

Does it make sense to conduct multiple analyses on textual data in order to check for robustness? And if it does, are such analyses likely to produce new knowledge concerning the meaning or interpretation of the corpora?

There is little in science that can be said to be known with absolute certainty. Usually, researchers must accept probabilities in lieu of certainties and so must ascertain at what level of probability one's results are likely to reflect the real world (or the larger population). In quantitative statistical analysis, accepted standards of probabilities are well-understood and fairly consistent (with higher levels expected in the hard sciences, where laboratory experiments allow better controls, and lower ones in the social sciences, where research is not subject to the same conditions). These probabilities help to demarcate strong findings from those that are more suspect.

We can also perform “checks” on our data, or in effect challenge our results in a variety of ways in order to see if our initial results are reasonably stable, or robust. So, if the

application of multiple content analysis software to the same textual data constitutes this type of robustness check, what is the appropriate number of checks to conduct—one, two, a dozen? There are no clear answers here. Moreover, even in quantitative statistical analysis, coefficients do not tell us about the intentions of actors or the meanings of outcomes: these are open to judgment. Automated content analysis of textual data poses an even greater challenge to researchers, because words are inherently subjective (users may assign different meanings to the same words, and differences in contexts may change meanings entirely). In short, meaning matters, and context matters. And yet, we seek to measure words and their meanings with some degree of statistical significance.

This paper has sought to assess the extent to which different automated content analysis software yield broadly similar results, when applied to the same corpora. And, the two corpora analysed derive from a lengthy book manuscript that seeks better to understand deliberations on US monetary policy. Given the extensive analysis completed in the book manuscript, we sought here to extract only our key findings for further analysis. Hence, the analysis here is only a small subset of the larger findings.

In order to come to some judgment regarding the reasonableness of conducting multiple analyses, we must address the second question first: having conducted multiple analyses, have we learned much more from our data? Did the added effort reveal new insights or important new findings? Well, not really. Our basic story concerning congressional oversight of monetary policy over the period from 1976-2008 was that (1) members of Congress did not engage with the Fed chairman on monetary policy per se, (2) there were marginal, but not fundamental, differences between the thematic focus of senators versus congressmen in their respective banking committees, and (3) there occurred a change over time in the apparent understanding of the contribution of sustained low inflation to US macroeconomic policy (especially a growing acceptance and lack of challenge to the low inflation consensus). Adding the analyses of two further automated content analysis software has not changed the substance of this story. From T-Lab, our analysis did raise the question of how best to measure (statistically) complex arguments that bridge disparate concepts and ideas. From Dtm-Vic, we were able to visualize the thematic time-line of the discourse more elegantly than in Alceste. But do either of these suggest that the added effort was worth it? Probably not—or at least, not if the reward is greater knowledge of the substance and meaning of the words under investigation. If, however, one construes the reward as one of greater certainty, then the answer is, well, yes, it *is* worth the added effort. We are more certain that our results and interpretation of the oversight hearings in the House and Senate banking committees are sound.

Chartier, J.-F. and J.-G. Meunier (2011) Text Mining Methods for Social Representation Analysis in Large Corpora. Papers on Social Representation (<http://www.psych.lse.ac.uk/psr/>) **20**, 1-37

Klüver, H. (2009). "Measuring Interest Group Influence Using Quantitative Text Analysis." European Union Politics **10**(4): 535-549.

Lahlou, S. (1995b). Vers une théorie de l'interprétation en analyse des données textuelles (Towards a theory of interpretation in text mining). 3rd International Conference on Statistical Analysis of Textual Data. S. Bolasco, L. Lebart and A. Salem. Rome: CISU, JADT 1995. **1**: 221-228.

Lahlou, S. (2011). "How Can We Capture the Subject's Perspective? An Evidence-Based Approach for the Social Scientist." Social Science Information **50**(3-4): 607-655.

Lancia, F. (2007). Word Co-occurrence and Similarity in Meaning: Some Methodological Issues (unpublished paper), <http://www.mytlab.com/wcsmeaning.pdf>: 39.

Lebart, L. (2003). Validation Techniques in Text Mining. Text Mining and its Applications. S. Sirmakessis, Springer: 169-178

Lebart, L., A. Morineau, et al. (1984). Multivariate Descriptive Statistical Analysis. New York, Wiley.

Lebart, L. and A. Salem (1994). Statistique Textuelle (Statistics for Text Mining). Paris, Dunod.

Lebart, L., A. Salem, et al. (1998). Exploring Textual Data. Dordrecht, The Netherlands, Kluwer Academic Publishers.

Lowe, W. (2003). Software for Content Analysis: A Review. Technical Report for the Identity Project, Weatherhead Center for International Affairs, Harvard University.

Lowe, W. and K. Benoit (2012). Qualitative Validation of Quantitative Text Scaling. 70th Annual Conference of the Midwest Political Science Association, Chicago.

Savaresi, S. M. and D. L. Boley (2001). On the Performance of Bisecting K-Means and PDDP. 1st SIAM Conference on Data Mining Chicago.

Schonhardt-Bailey, C. (2013). Deliberating American Monetary Policy: A Textual Analysis. Cambridge, MA, MIT Press.

Schonhardt-Bailey, C., E. Yager, et al. (2012). "Yes, Ronald Reagan's Rhetoric was Unique—But Statistically, How Unique?" Presidential Studies Quarterly(September).

Steinbach, M., G. Karypis, et al. (2000). A Comparison of Document Clustering Techniques. Proceedings of the World Text Mining Conference, Boston.

¹ Iramuteq (<http://www.iramuteq.org/>) reproduces the Alceste method and is freely available. It is based on the R statistical software (<http://www.r-project.org/>) and written in Python language (<http://www.python.org/>). While Iramuteq reproduces the Alceste double classification, its interface is currently only in French (although there are promises of a future English version).

² Both T-Lab and Dtm-Vic provide extensive manuals and documentation on-line (see <http://www.tlab.it/en/presentation.php> and http://www.dtmvic.com/06_ManualE.html). A fuller description of Alceste is available in our book appendix I, and further documentation on Alceste and text analysis may be found on the links page of my website (<http://personal.lse.ac.uk/schonhar/>).

³ The default χ^2 threshold for selection of characteristic statements (ECUs) is 0, and for tags it is 2. ECUs with χ^2 values below 0 are unclassified; hence, the percent of classified ECUs constitutes a goodness of fit measure.

⁴ See (Lahlou 1995b) for a detailed description of the interpretation procedure and its theoretical basis.

⁵ We are very grateful to Ludovic Lebart, who kindly answered many questions relating to Dtm-Vic, particularly with regard to interpretation of results, relative to those of Alceste and T-Lab.

Table 1: ALCESTE Basic Statistics for House and Senate Hearings on Monetary Policy (1976-2008)

	House Hearings, 1976-2008	Senate Hearings, 1976-2008
Total Word Count (number of retained occurrences)	758,092	700,368
Unique Words Analyzed (i.e., occurring with frequency > 3)	312,071	297,108
Final Minimum Frequency of an Analyzed Word	33	31
Meaningful Words	1380	1396
Passive Variables (year, role, party, name—where each name constitutes a unique variable)	237	129
I.C.U.s (= number of speeches / comments)	6,237	5,744
Classified E.C.U.s	7,506 (= 92% of the retained E.C.U.)	6,984 (= 91% of the retained E.C.U.)
Lexical Classes	9	9
Distribution of Classes (%) and Thematic Content	1 (14) <i>Populist Attack on Fed/Greenspan (B. Sanders)</i> 2 (13) <i>Volcker Defending Anti-Inflation Stance (give & take; speculative)</i> 3 (13) <i>Fiscal Policy</i> 4 (12) <i>Fed's Regulatory Activity</i> 5 (12) <i>Q & A Format (Process); Mixed Substance</i> 6 (11) <i>Monetary Aggregates</i> 7 (10) <i>US Real Economy</i> 8 (9) <i>MCs Prompting Fed Chair on Non-Monetary Issues</i> 9 (6) <i>Capital Inflows, Exchange Rate, Current Account Deficit</i>	1 (14) <i>World Economy & US External Balance (Trade & Current Account)</i> 2 (13) <i>Bank Regulation & Banking Industry Structure</i> 3 (12) <i>Q & A Format (Volcker trying to define limits of Fed's knowledge / role)</i> 4 (12) <i>Fed Appointments & Relationship between Fed, Congress & Administration</i> 5 (9) <i>Education, Training & US Competitiveness (Labour Market)</i> 6 (9) <i>Fiscal Policy</i> 7 (9) <i>Monetary Aggregates & Objectives of Monetary Policy</i> 8 (9) <i>US Real Economy</i> 9 (14) <i>Criticism of Fed for failing to support growth (D. Riegle)</i>

Table 2: T-LAB Basic Statistics for House and Senate Hearings on Monetary Policy (1976-2008)

	House Hearings, 1976-2008	Senate Hearings, 1976-2008
Total Word Count	731,841	680,488
Single Words or Multi-words (“strings” analysed)	15,789	14,912
Threshold Frequency for Words	11	12
Key Terms	980	980
Variables (year, role, party, name—where name is dropped from analysis as N > 150)	4	4
Primary Documents [texts] (= number of speeches / comments)	6,237	5,744
Classified Elementary Contexts (in “chunks”)	16,304 (95%)	15,322 (95%)
Clusters Available	3 to 10 (default = 3)	3 to 10 (default = 4)
Distribution of Classes (%) and Thematic Content	1 (15) <i>Inflation & Prices</i> 2 (11) <i>Banking Regulation</i> 3 (9) <i>Bank Lending & Credit Creation</i> 4 (14) <i>Uncertainty</i> 5 (9) <i>Monetary Aggregates</i> 6 (15) <i>Fiscal Policy</i> 7 (12) <i>Monetary Policy</i> 8 (9) <i>Fed/Congress: Committees</i> 9 (7) <i>Labour Market</i>	1 (14) <i>Real Economy & Labour Market</i> 2 (12) <i>Inflation & Prices</i> 3 (15) <i>Change in Variables (relating to the real economy)</i> 4 (8) <i>Financial Regulation: Fed’s Role</i> 5 (7) <i>Mix between Fiscal & Monetary (role of Congress & Fed)</i> 6 (15) <i>Monetary Aggregates</i> 7 (10) <i>Fiscal Policy</i> 8 (11) <i>Banking Structure & Regulation</i> 9 (9) <i>Questioning the Direction of Monetary Policy</i>

Table 3: Dtm-Vic Basic Statistics for House and Senate Hearings on Monetary Policy (1976-2008)

	House Hearings, 1976-2008	Senate Hearings, 1976-2008
Total Word Count (Word Count after Lemmatization, using TreeTagger)	765,586 (459,432)	708,173 (463,085)
Distinct Words (Distinct Words after Lemmatization)	19,268 (11,459)	17,850 (11,471)
Frequency Threshold	50	200
Words retained following Frequency Test (Distinct Words retained)	397,728 (1242)	323,489 (441)
Passive Variable (year)	1	1
Number of speeches / comments	6,237	5,744
Thematic Content of Year Variable (20 periods):		
1976	Monetary Aggregates (& Labour Market)	Monetary Aggregates (& Labour Market)
1977	Supply-Side (Businesses & Labour Market)	Unemployment/Businesses/Inflation
1979	Monetary Aggregates (& Inflation)	Monetary Aggregates (& Credit Creation)
1980	Credit Control (& Inflation)	Credit Control (& Inflation)
1981	Interest Rates	Money & Interest Rates
1984	Fiscal Policy (& Interest Rates)	Bank Lending/Dollar
1985	Monetary Aggregates (Focus on the Dollar)	Current Acct Deficit/Monetary Aggregates
1986	Banks	International Focus/Banks/Monetary Policy
1991	Currency/Financial/Greenspan/Inflation	Banks/Capital Reserves
1992	Fiscal Policy	Labour Market/Recession/Greenspan
1993	Fiscal Policy	Small Businesses/Fiscal Policy/Deficit Reduction
1997	Labour Markets	Interest Rates/Fed's Reserve Holdings/Deposits
1998	US Economy & Productivity	US Economy & Productivity
1999	US Economy & Trade	Banks/Financial Services
2003	Labour Market & Fiscal Policy	(Process related) Greenspan/FOMC
2004	Labour Market	Labour Market/Productivity/Greenspan
2005	Social Security/Greenspan	Social Security (?)
2006	Labour Market/Housing/Inflation	Energy Prices/Inflation/Housing
2007	Housing (Mortgages)	Housing (Mortgages)
2008	Housing/Credit Crisis/Risk	Housing/Mortgages/Risk

Table 4: Common Themes Between Alceste and T-Lab

COMMON THEMES	HOUSE THEMATIC CLASSES (%): ALCESTE	HOUSE THEMATIC CLASSES (%): T-LAB	SENATE THEMATIC CLASSES (%): ALCESTE	SENATE THEMATIC CLASSES (%): T-LAB
Fiscal Policy	Fiscal Policy (13)	Fiscal Policy (15)	Fiscal Policy (9)	Fiscal Policy (10)
Regulation	Fed's Regulatory Activity (12)	Banking Regulation (11)	Bank Regulation & Banking Industry Structure (13)	Financial Regulation (Fed's Role) (8) ... Banking Structure & Regulation (11)
Monetary Aggregates	Monetary Aggregates (11)	Monetary Aggregates (9)	Monetary Aggregates & Objectives of Monetary Policy (9)	Monetary Aggregates (15)
Real Economy	<i>US Real Economy(10)</i>	<i>Monetary Policy (12)</i>	US Real Economy (9)	Real Economy & Labour Market (14) ... Change in Variables (relating to the Real Economy) (15)
Uncertainty & Challenges	<i>Volcker Defending Anti-Inflation Stance (give & take; speculative) (13)</i>	<i>Inflation & Prices (15)</i> ... <i>Uncertainty (14)</i>	<i>Q & A Format (Volcker trying to define limits of Fed's knowledge / role) (12)</i>	<i>Questioning the Direction of Monetary Policy (9)</i>
			<i>Criticism of Fed for failing to support growth (D. Riegle) (14)</i>	<i>Questioning the Direction of Monetary Policy</i>
	<i>Populist Attack on Fed/Greenspan (B. Sanders) (14)</i>	<i>Labour Market (7)</i>		
Process & Fed/Congress	<i>Q & A Format (Process); Mixed Substance (12)</i>	<i>Fed/Congress: Committees (9)</i>	<i>Fed Appointments & Relationship btwn Fed, Congress & Administration (12)</i>	<i>Mix between Fiscal & Monetary (role of Congress & Fed) (7)</i>
OTHER CLASSES & CLUSTERS				
MCs Prompting Fed Chair on Non-Monetary Issues (9)				
Capital Inflows, Exchange Rate, Current Account Deficit (6)				
		Bank Lending & Credit Creation (9)		Inflation & Prices (12)
			World Economy & US External Balance (Trade & Current Account) (14)	
			Education, Training & US Competitiveness (Labour Market) (9)	

Figure 1: Summary Results from *Deliberating Monetary Policy*

Figure 2a: Tree Diagram of Relative Associations between Classes for House Hearing, 1976-2008

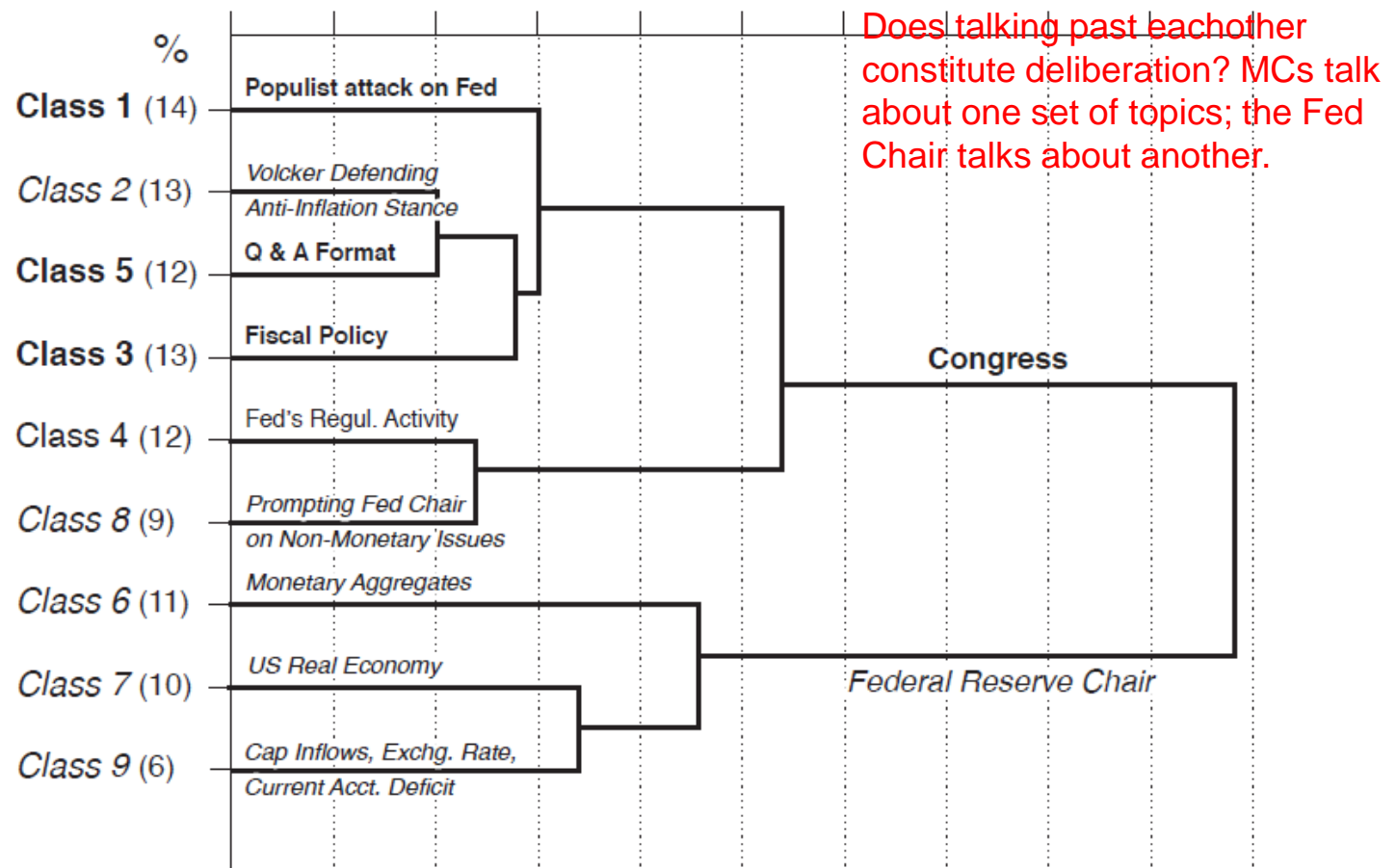


Figure 2: Summary Results from *Deliberating Monetary Policy*

Figure 2b: Tree Diagram of Relative Associations between
Classes for Senate Hearing, 1976-2008

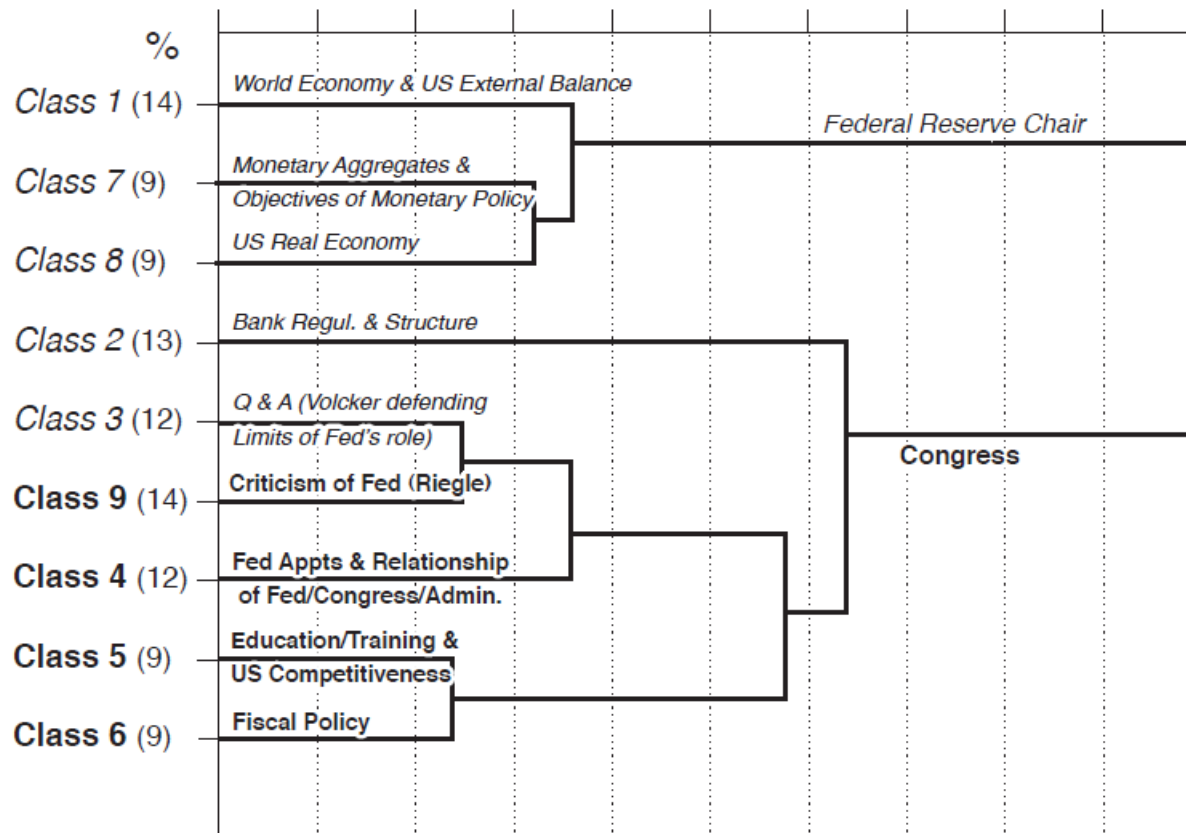
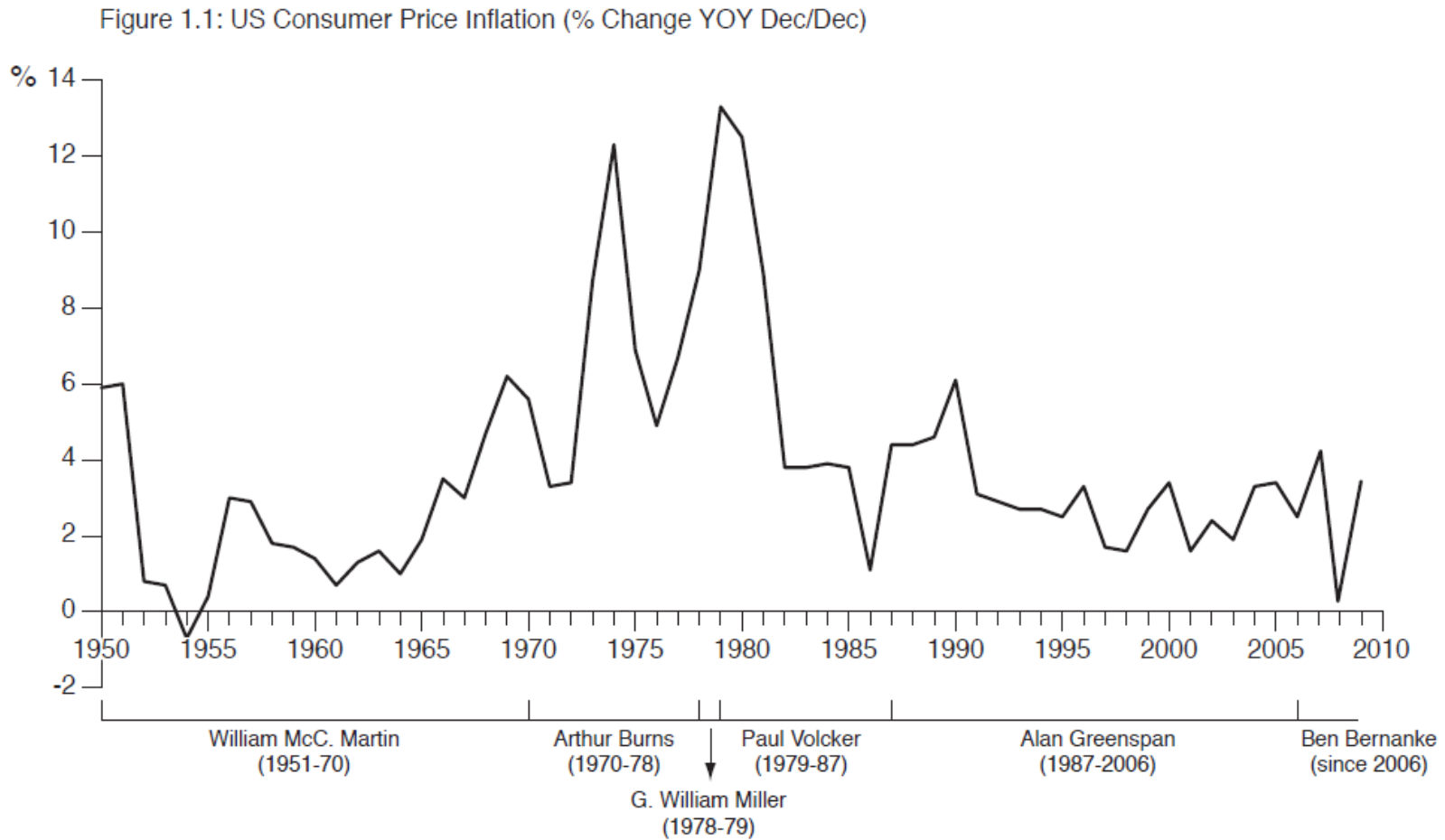


Figure 3: Rate of Inflation over Time



Source: US Department of Labor, Bureau of Labor Statistics

Figure 4: Dendrogram of T-Lab Clusters for House Hearings 1976-2008

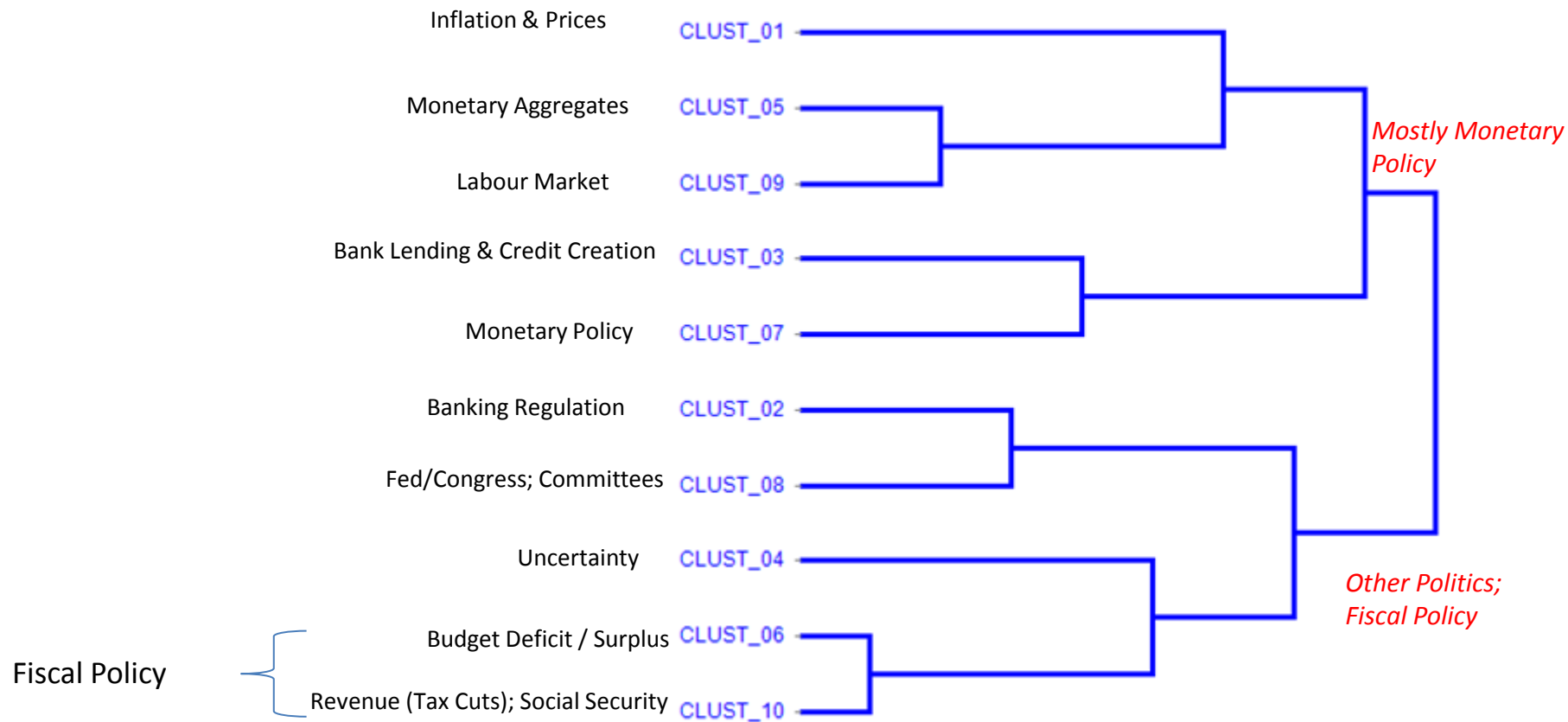


Figure 5: Dendrogram of T-Lab Clusters for Senate Hearings 1976-2008

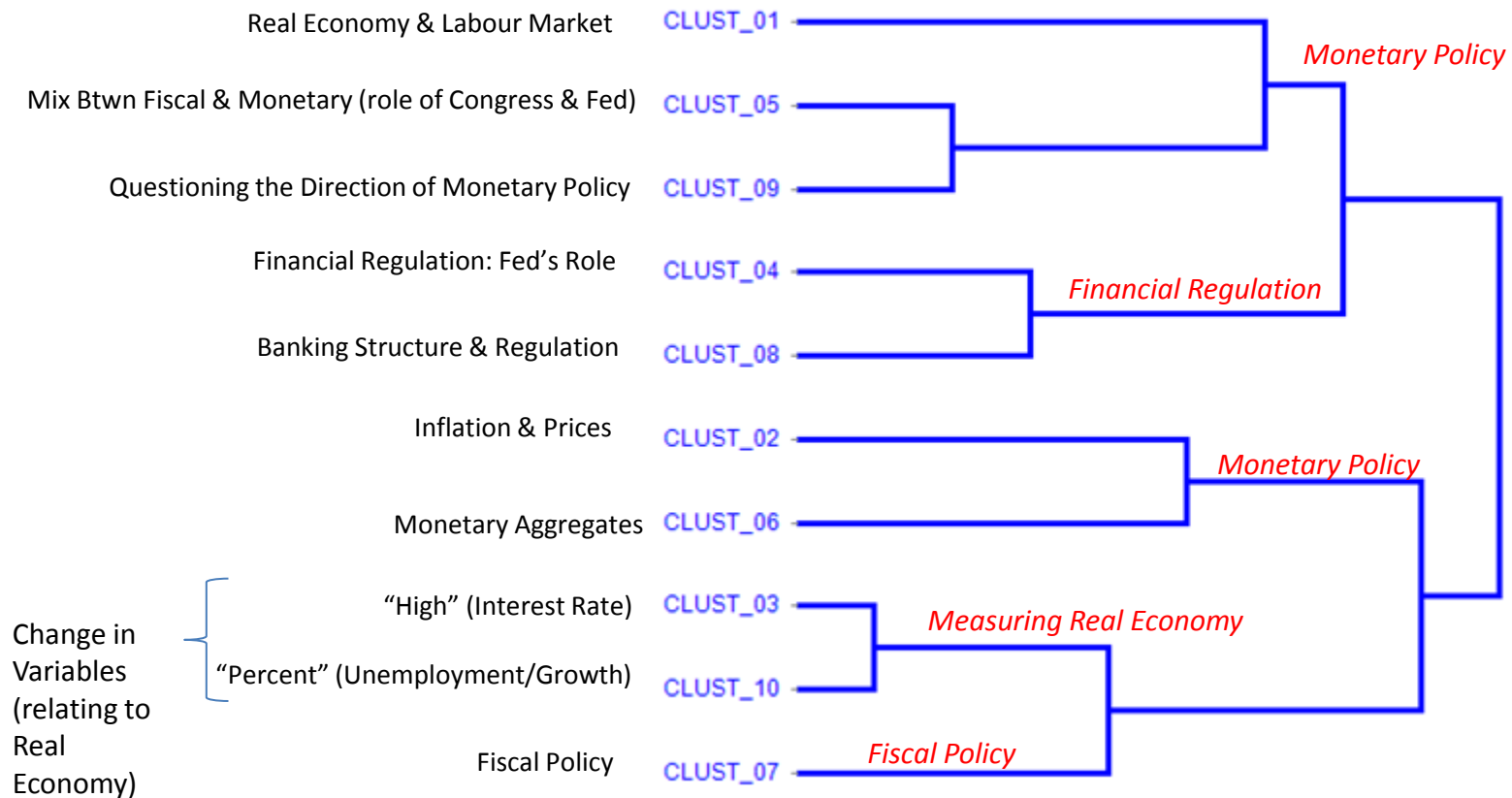


Figure 6: House Hearings, 9 Cluster Partition, by Role of Participant – Top 3 Clusters

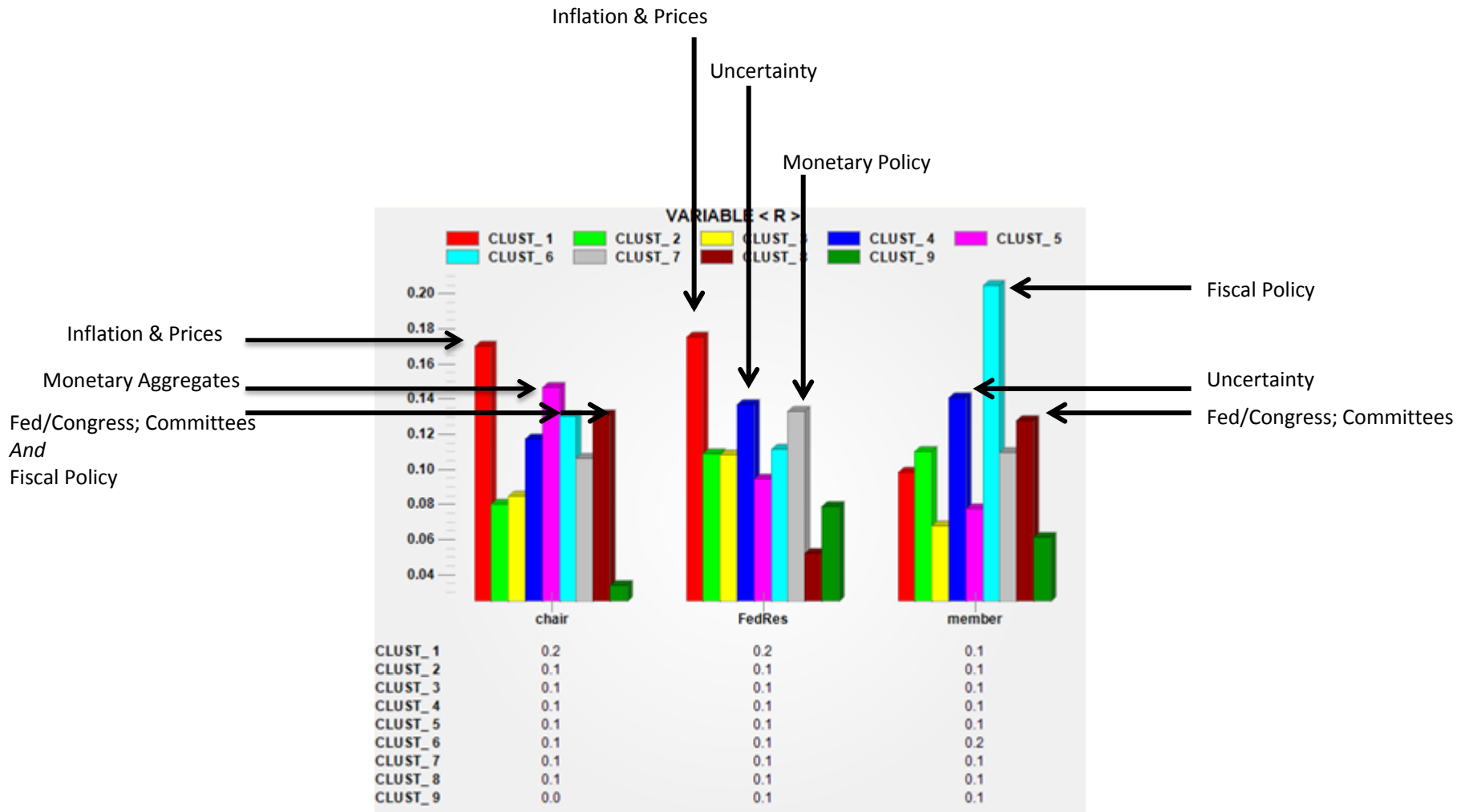


Figure 7: Senate Hearings, 9 Cluster Partition, by Role of Participant –
Top 3 Clusters

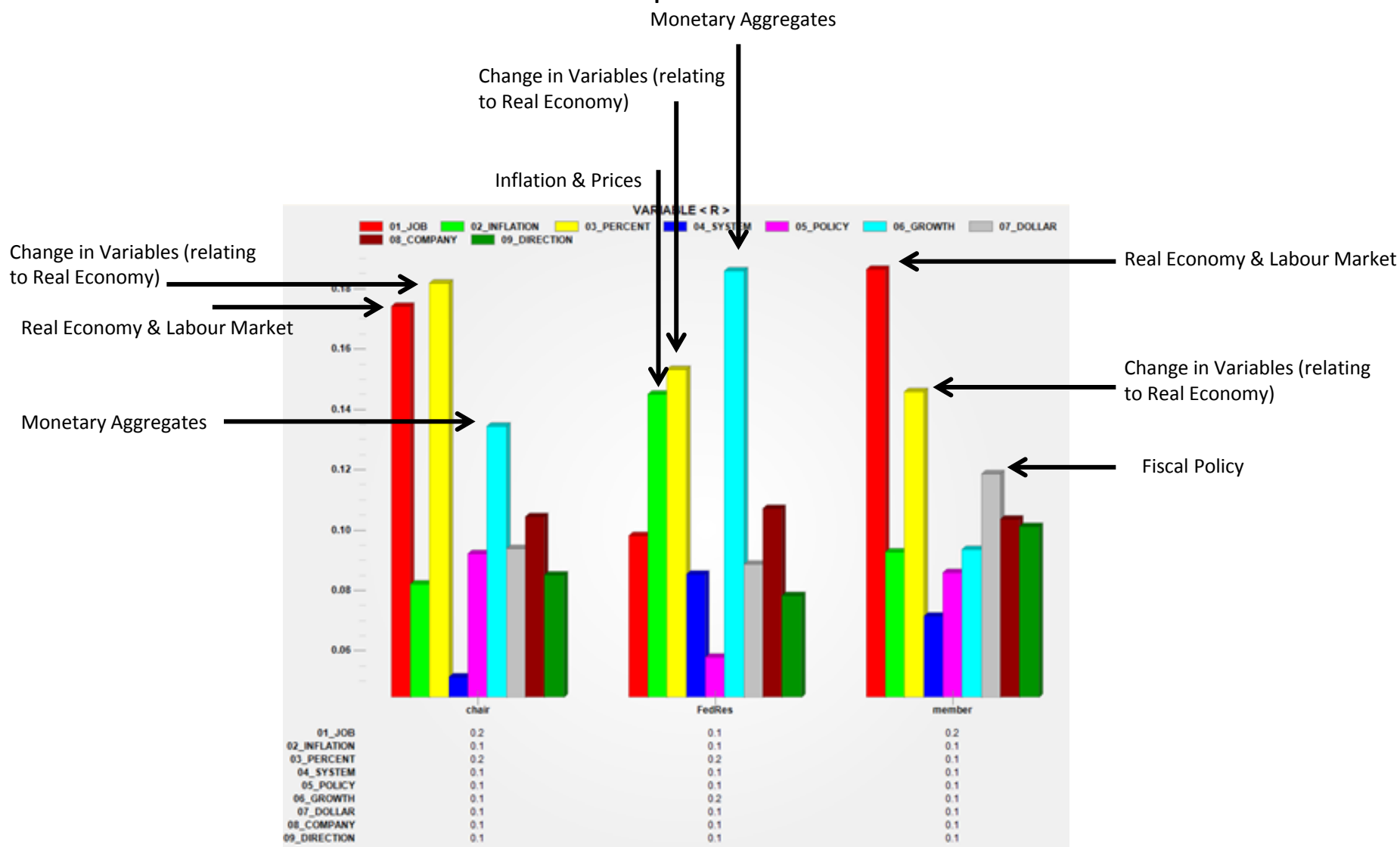


Figure 8: Senate Hearings, 9 Cluster Partition, by Party Affiliation

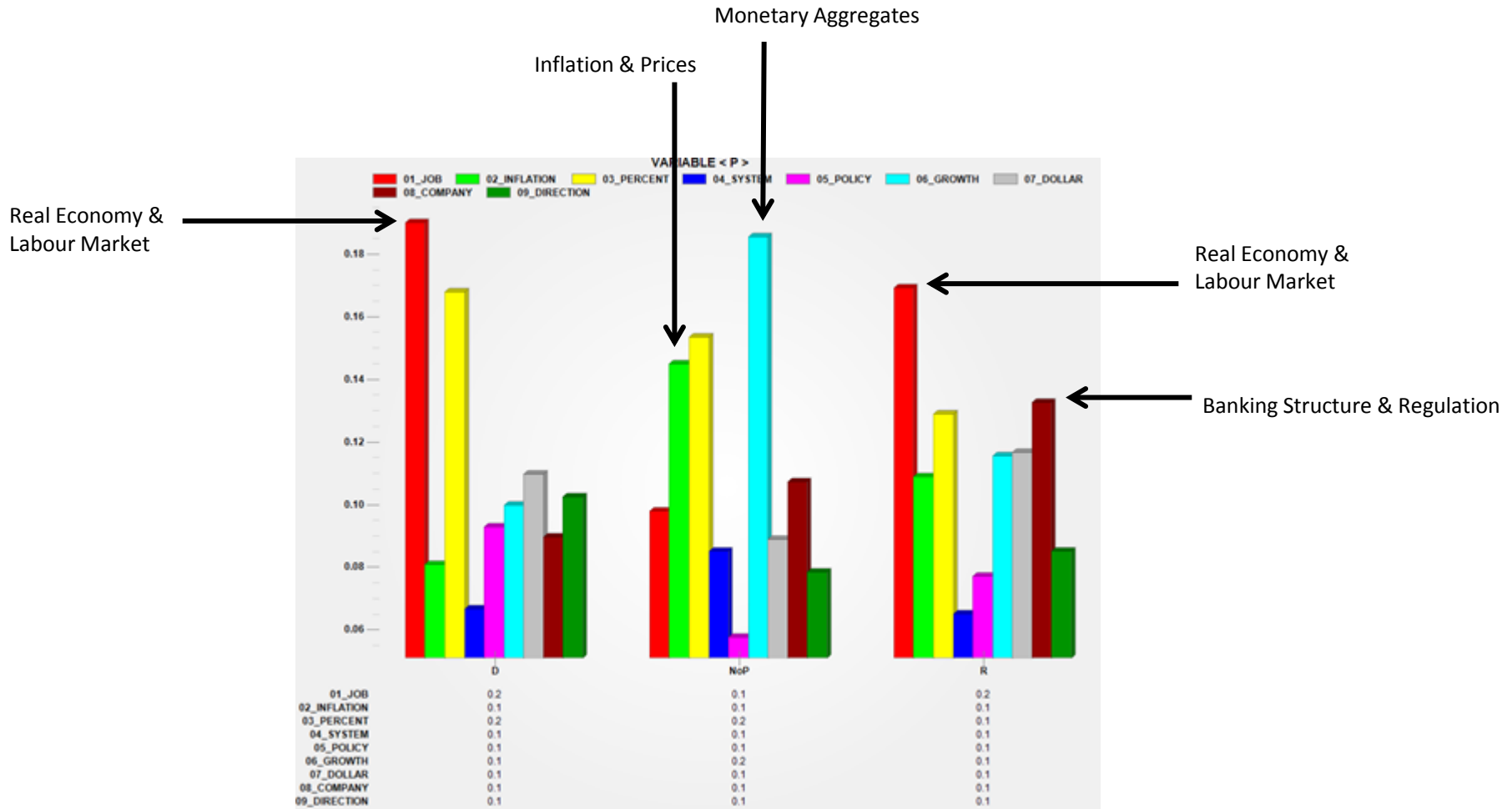
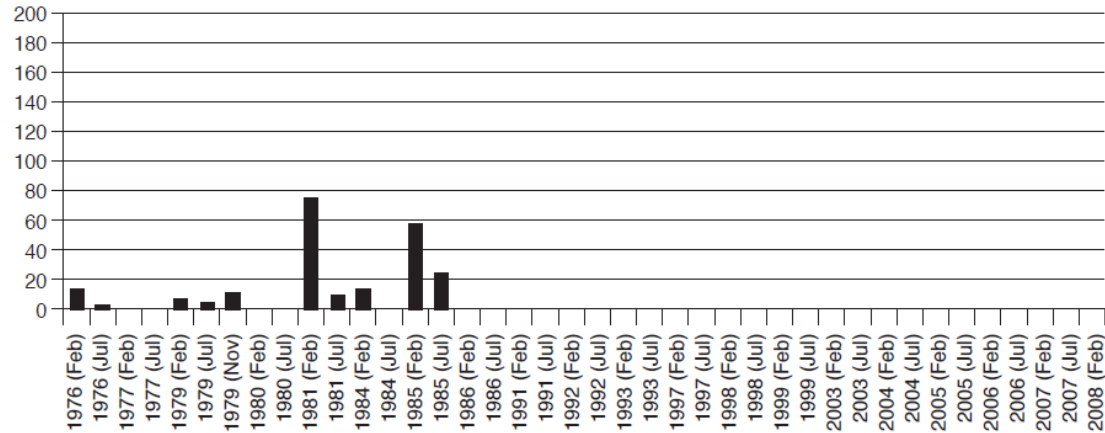


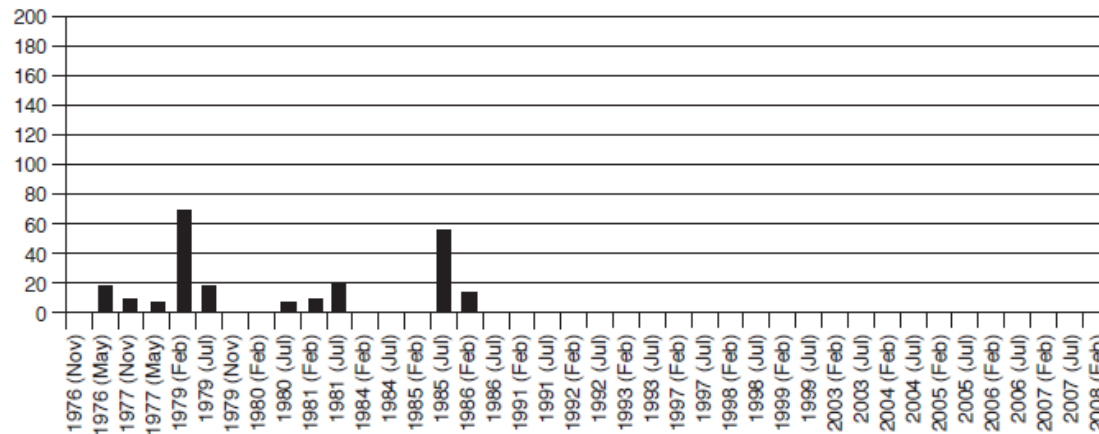
Figure 9: Alceste results for Monetary Aggregates theme

Figure 4.A6: House Hearings 1976-2008 – Class 6: Monetary Aggregates
[Chi-square values]



House

Figure 4.A16: Senate Hearings 1976-2008 – Class 7: Monetary Aggregates
[Chi-square values]



Senate



Exploring Numerical and
Textual Data in Practice
with Dtm-Vic

First table analysed
through
Correspondence
Analysis:

Contingency table
cross-tabulating 20
years and the 1242
lemmas appearing at
least 50 times.

Figure 10: House Hearings: Chronology of themes in corpus

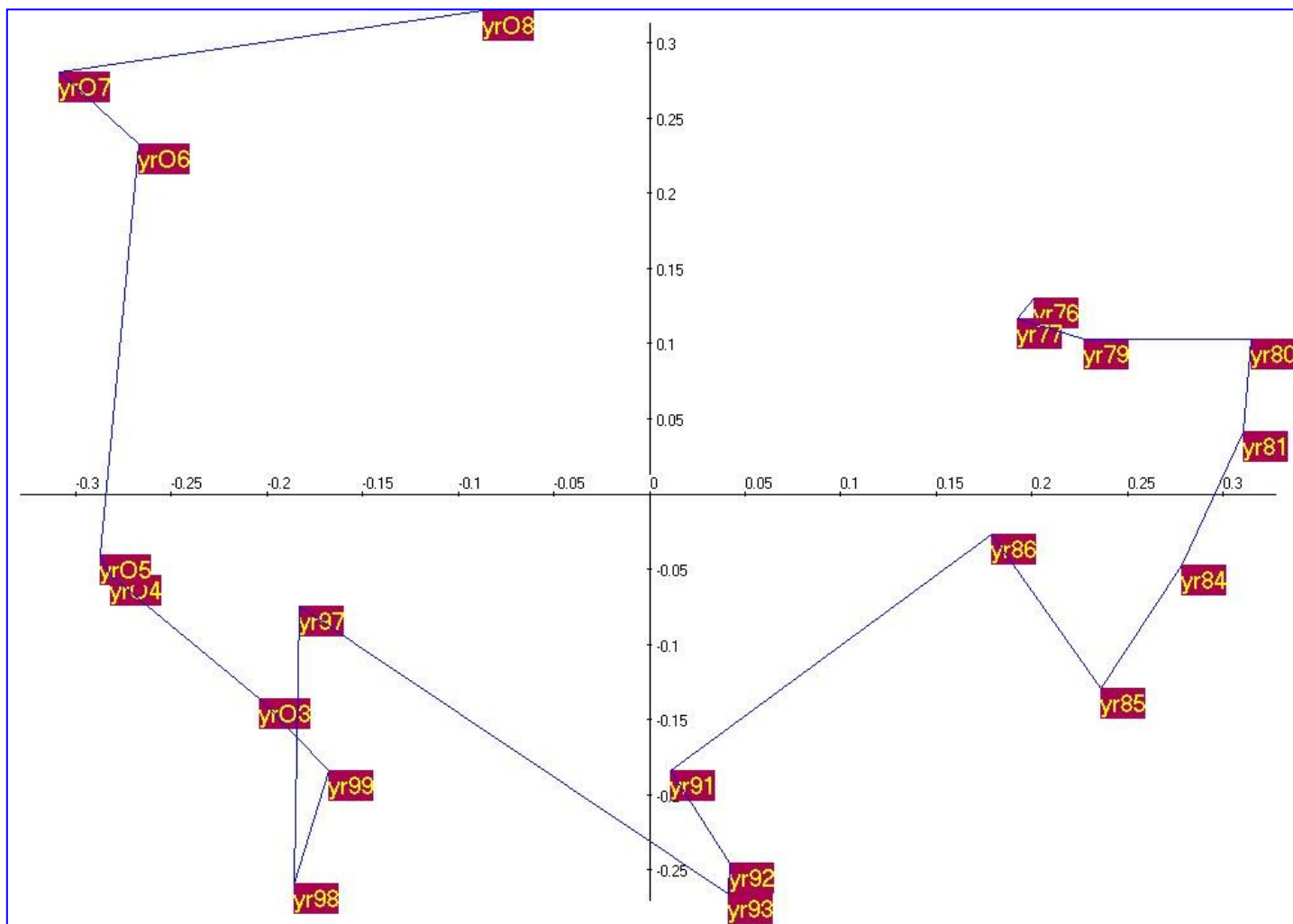


Figure 11: Second table analysed through Correspondence Analysis (testing a very large threshold of frequency)

Contingency table cross-tabulating 20 years and the 303 lemmas appearing at least 300 times.

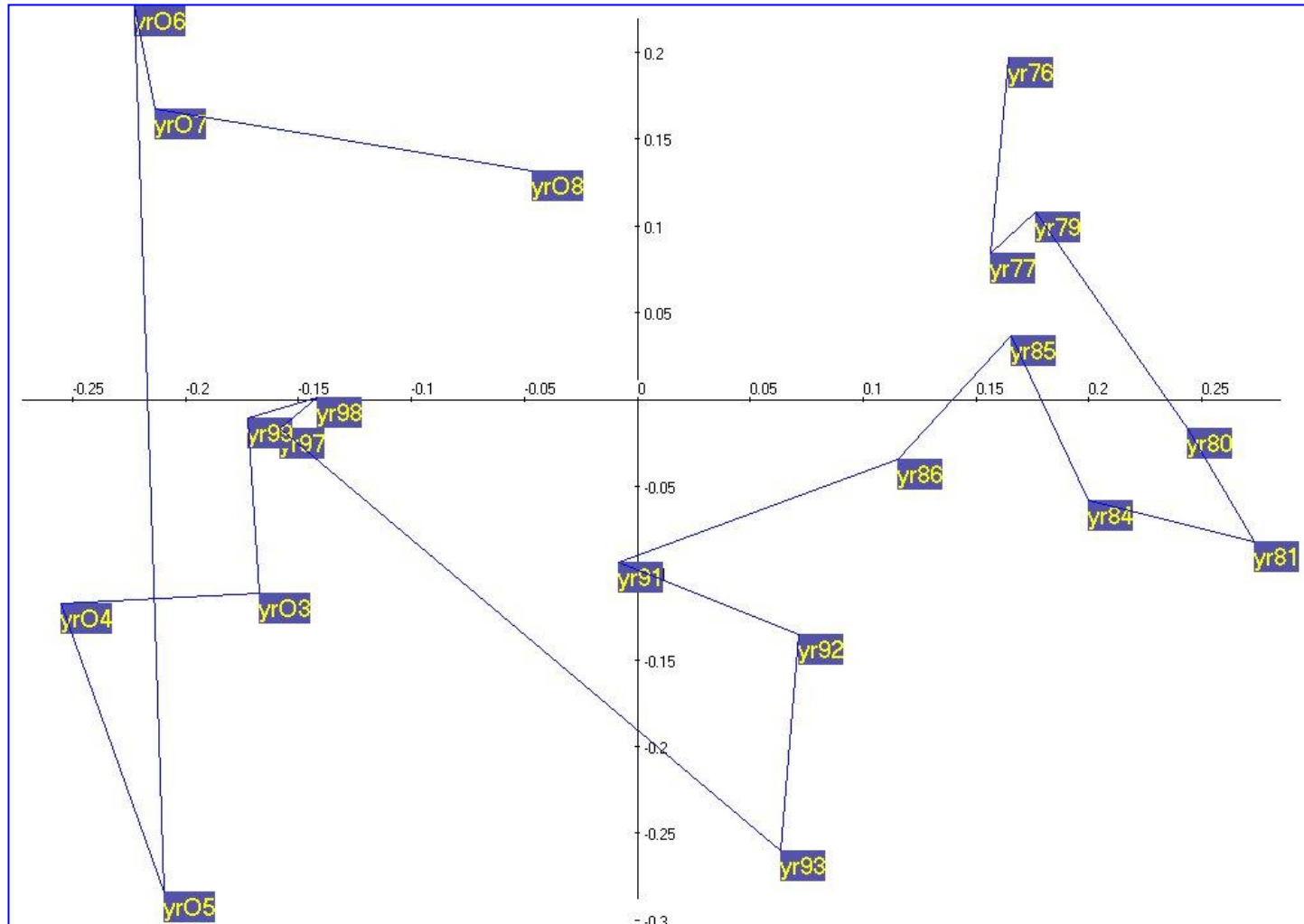


Figure 12: House Hearings (using very high frequency threshold) - 6
Clusters of contiguous years that are homogeneous from a lexical point of view: (1) late 1970s; (2) mid-1980s; (3) early 1990s; (4) late 1990s; (5)
2003-05; and (6) 2006-08

Early period in right quadrants; later period in left quadrants

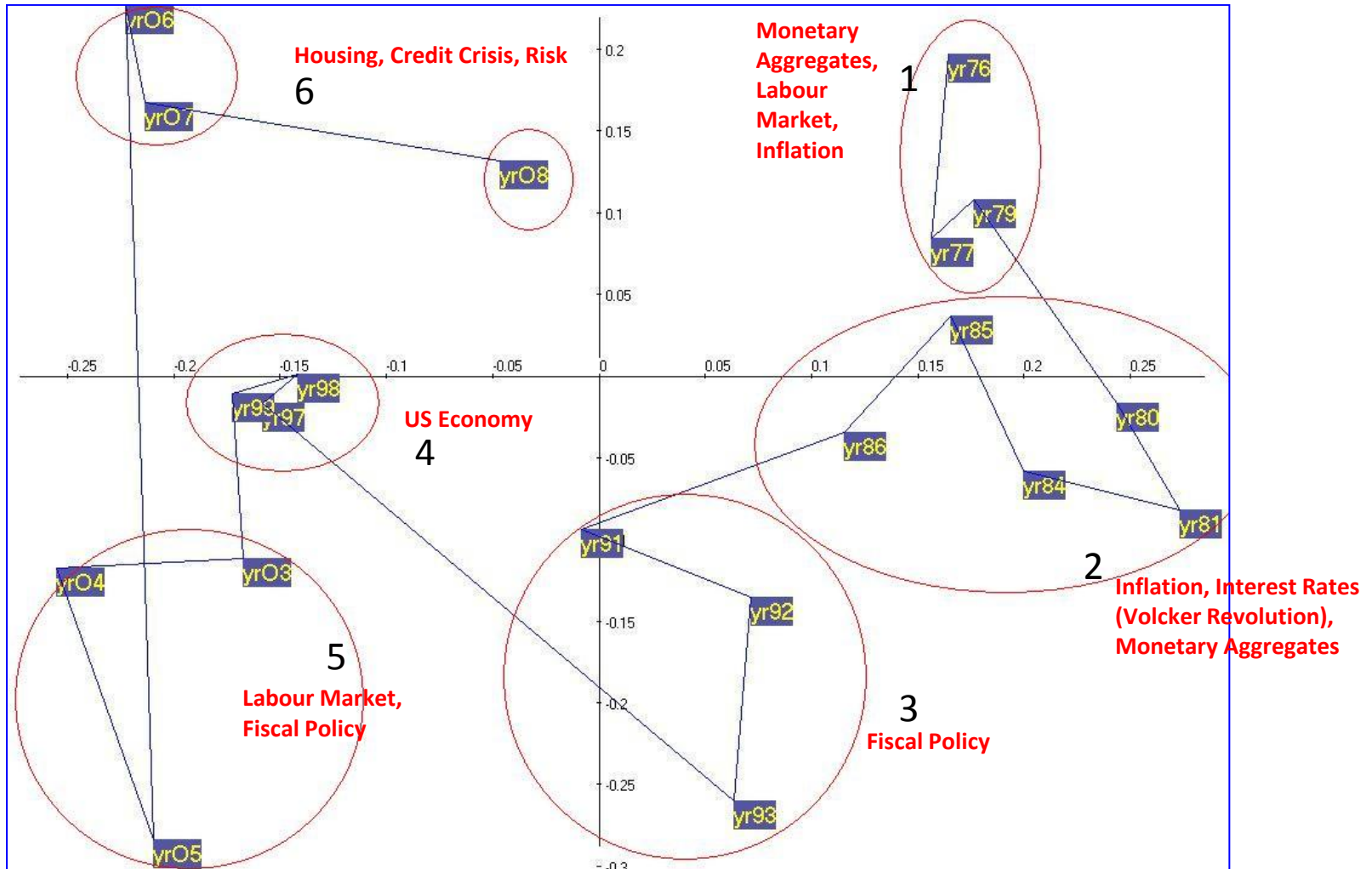


Figure 13: SENATE HEARINGS - First table analysed through Correspondence Analysis:
 Contingency table cross-tabulating 20 years and the 441 lemmas appearing at least 200 times.

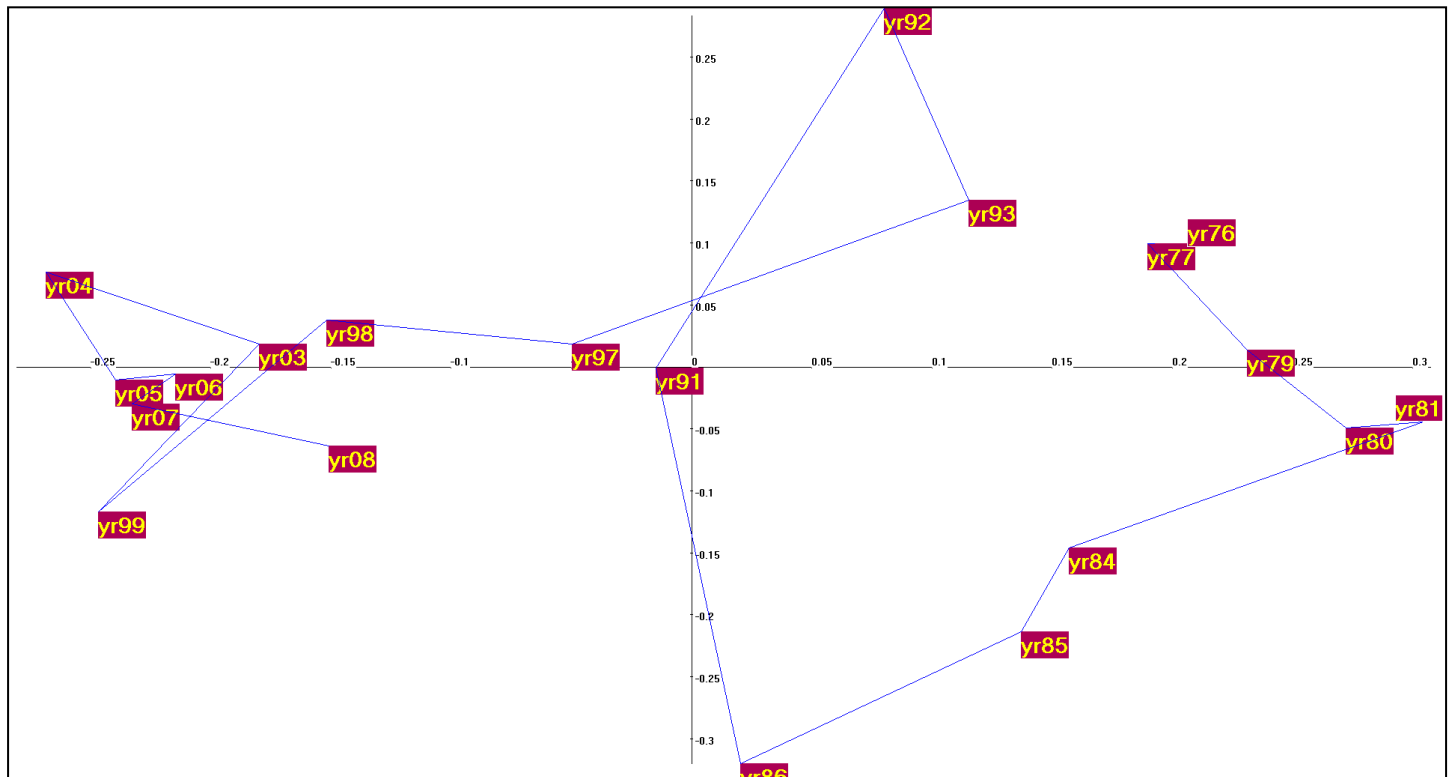
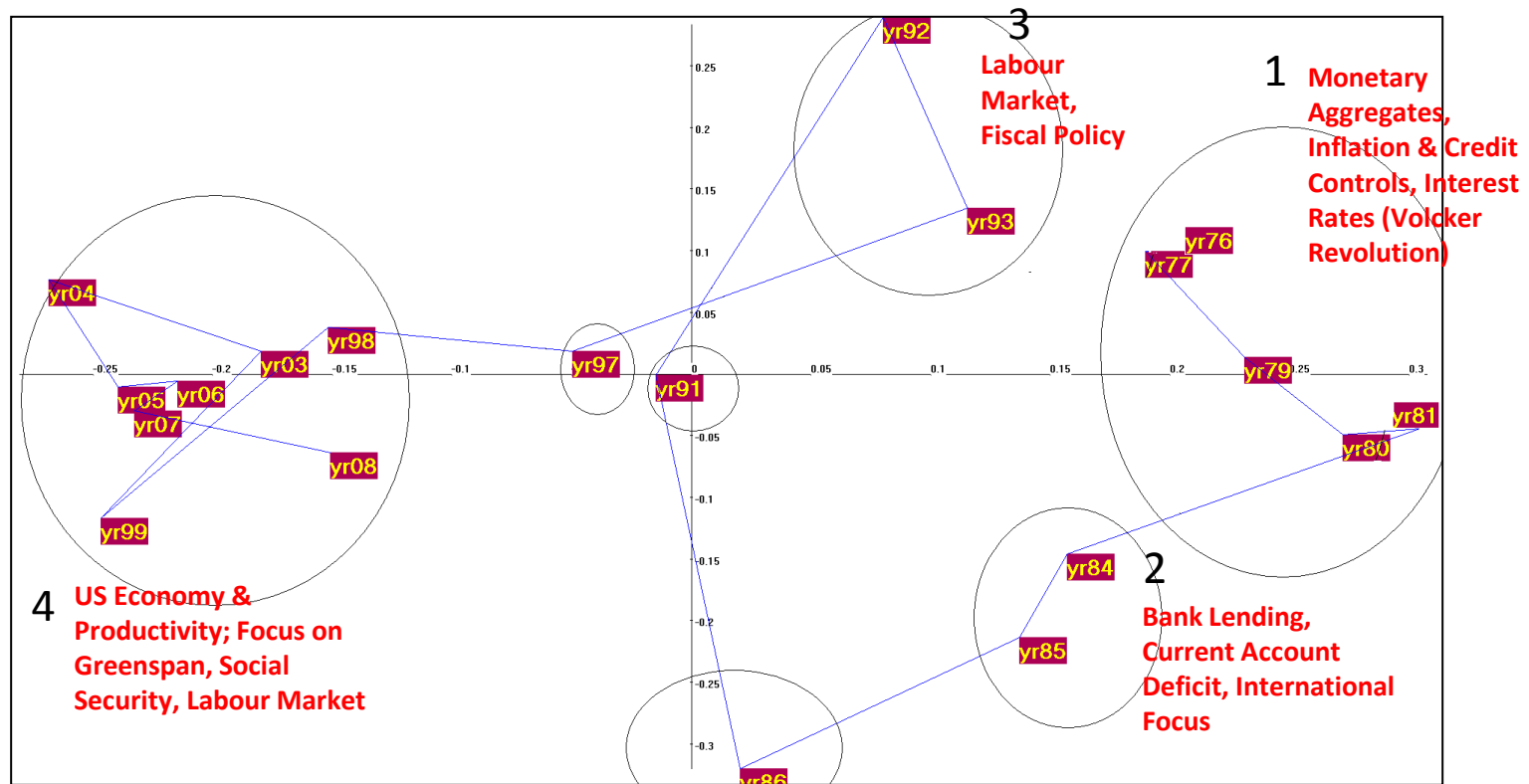


Figure 14: SENATE HEARINGS - 4 Clusters of contiguous years that are homogeneous from a lexical point of view: (1) late 70s/early 80s; (2) mid-80s; (3) early 90s; (4) late 90s & early 2000s

Early period in right quadrants; later period in left quadrants

(Three years remain isolated: 1986, 1991, 1997.)



**Figure 15: Example of statistical inference on textual data (Senate corpus).
Confidence areas in the plane spanned by axes 1 and 2 from the
correspondence analysis of the lexical table (years x words)**

