

The Cameelious Hump and 'Just So Stories' in Economic Literature – A Linguistic Analysis

Benedikt Szmrecsanyi and Nils Goldschmidt

Introduction

*I get it as well as you-oo-oo--
If I haven't enough to do-oo-oo--
We all get hump--
Cameelious hump--
Kiddies and grown-ups too!*

Rudyard Kipling, *Just So Stories*

More than a century ago, the British writer Rudyard Kipling published his *Just So Stories* (Kipling, 2002 [1902]). In this collection, Kipling offers some famously imaginative and affectionate accounts of a number of the most wondrous things in the world: How the Whale got his Throat, How the Camel got his Hump, How the Rhinoceros got his Skin, How the Leopard got his Spots – just to name the first four stories. Our point is that Kipling playfully engaged in an exercise that would be labeled 'abduction' in the parlance of the pragmatist philosopher Charles Sanders Pierce – that is, generating a hypothesis that suggests that something *may be*. In criticizing Richard Posner's law-and-economics approach, Jeanne L. Schroeder summarizes Kipling's original motivation for his 'just so stories' as follows:

Kipling implicitly understood that although "abduction" – the post hoc imagining of explanatory stories – can play an essential role in scientific inquiry, it is merely a means of forming, not proving, hypotheses. One can easily abduct explanations that, while sounding plausible within their context, appear fantastic when examined more fully. Consequently, Kipling created his just so stories as fairy tales for the amusement of his daughter. In contrast, Posner claims his just so stories are scientific theories for the edification of the legal community. He wants his audience to accept his economic analysis not merely as an account of past legal developments, but as a model for future ones. (Schroeder, 2001: 351-352)

While Schroeder, drawing on Kipling, took issue with the law-and-economics approach specifically, McCloskey (1985) – more than twenty years ago – had already begun a similar endeavor, albeit one taking on economic science in its entirety. McCloskey pointed out that economists are working neither deductively nor inductively, really, but that – and let us adhere to Pierce's terminology now – they are actually proceeding 'abductively', telling mere 'just so stories': "Pure theory in economics is similar to the literary of fantasy. Like fantasy it violates the rules of 'reality' for convenience of the tale; and, of course, amazing results become commonplace in a world of hypotheses" (McCloskey, 1991: 70). In other words, a close reading of both Kipling and McCloskey should make one wonder about some economic tales: How the market got his invisible hand, How competition found its equilibrium, How natural man became a homo economicus, and so on.

Thus given that, according to McCloskey, economists ultimately recite post-hoc stories, we are left to wonder *how* these stories are told, whether economic fairy tales are narrated *differently* than, say, sociological fairy tales, and *what* economic fairy tales are about. These

questions are the ones that will take center stage in our paper. We will draw on linguistic methods to explore large-scale patterns of variance in the textual material published in economic journals, as well as in journals covering other professions, and in doing so we embark on an empirical investigation of scholarly (and particularly economic) story-telling. We believe that our methodology can shed light on the question which stories are favored by economists, and what these stories reveal about contemporary orthodox economic theory.

Data

As for our empirical database, we compiled a linguistic corpus – which is a technical term for a linguistic database of authentic texts in machine readable form which represents a sample of a particular linguistic variety or genre – designed to facilitate (1) a comparison between economic academic prose and academic prose in other disciplines, on the one hand, and (2) an analysis of shifting emphases in economic journals – now and over time – on the other hand. Our corpus includes samples from seven journals (four of which are economic) and, for economics, the period from the mid-1960s until today, with sample times spanning 1965, 1980, and the 1990s. The following journal samples were included in our corpus to represent contemporary¹ economic scholarly writing:

- The *American Economic Review* (henceforth: AER). A leading, top-ranked economic journal in the Social Citation Index, AER is one of the flagship publications of the discipline and prides itself of its high scientific standards and its focus on highly original research. See www.aeaweb.org/aer/ for the journal's web site. We included Vols. 80–89 (1990–1999) Issue 3 each (sample size: c. 1.94 million words) as well as Vol. 55 Issue 3 (1965; sample size: 0.17m words) and Vol. 70 Issue 3 (1980; sample size: 0.12m words).
- The *Economic Journal* (henceforth: EJ). Dedicated to "promoting the advancement of economic knowledge" (www.jstor.org/journals/00130133.html), EJ "is a key source for professional economists in higher education, business, government service and the financial sector, and represents unbeatable value for those who want to keep abreast in current thinking in economics" (www.blackwellpublishing.com/journal.asp?ref=0013-0133). EJ does not publish original research only but also pieces intended for a wider audience and devoted to the reflection of newer, theoretical developments in economics. We sampled Vol. 108 (1998) Issues 446–451; Vol. 109 (1999) Issues 452–459; and Vol. 110 (2000) Issues 460–465 (sample size: 2.42m words) as well as Vol. 75 Issues 298 and 299 (1965; sample size: 0.17m words) and Vol. 90 Issues 358 and 359 (1980; sample size: 0.28m words).
- The *Journal of Economic Perspectives* (henceforth: JoEP) "attempts to fill a gap between the general interest press and most other academic economics journals", as the publisher asserts (see www.aeaweb.org/jep/). JoEP does not feature original research but, for the main part, summarizes and comments on earlier work. Included in our sample are Vols. 4–13 (1990–1999) Issues 2 and 3 each. This sample has a size of 1.90m words.
- The *American Journal of Economics and Sociology* (henceforth: AJES). This journal "sets no ideological standards for its collaborators or contributors" (Moss 2002, 1): "Today, the exciting and unrelenting encounters between sociology and economics are a natural subject to explore, and AJES continues to publish carefully crafted essays in the social

¹ For simplicity, we will refer to texts dating from the 1990s and later as 'contemporary' texts in the remainder of this study.

sciences that represent interesting analysis with an eye toward assisting both practitioners and academicians to better understand the world in which we live." (<http://www.blackwellpublishing.com/journal.asp?ref=0002-9246>). We chose to include AJES in our analysis since it is, we think, at the forefront when it comes to accommodating sociological issues in economics. We included Vols. 49–60 (1990–2001) Issues 2, 3, and 4 each (sample size: 2.11m words) as well as Vol. 24 Issues 2, 3, and 4 (1965; sample size: 0.11m words) and Vol. 39 Issues 2, 3, and 4 (1980; sample size: 0.10m words).

In addition, the following samples, drawn from leading journals of other disciplines – sociology, microscopy, and mathematics – were included for benchmarking purposes. As such, these data samples will serve to triangulate economic scholarly prose within a threefold space of rather socio-scientific academic prose (sociology), rather applied natural science academic prose (microscopy), and rather formal-abstractive academic prose (mathematics).

- The *American Journal of Sociology* (henceforth: AJS). AJS was established in 1885 as the first US journal in its field and is the "leading voice for analysis and research in the social sciences, presenting work on the theory, methods, practice, and history of sociology", as the publisher asserts (see www.journals.uchicago.edu/AJS/brief.html). We sampled Volume 106 Issue 3 (2000; size: 0.13m words).
- The *Journal of Microscopy* (henceforth: JOM). In our study, we will take JOM to be a typical representative of writing in the applied natural sciences. It covers "all aspects of microscopy and high-energy *in situ* beam analysis" as well as "the application of microscopical techniques or specimen preparation procedures in an investigation" (see www.blackwellpublishing.com/journals/jmi/). We included Vol. 206 Issues 1, 2, 3 (2000) and Vol. 207 Issue 1 (2000). This sample has a size of 0.18m words.
- The *Journal of the American Mathematical Society* (henceforth: JAMS). JAMS is a renowned mathematical magazine "devoted to research articles of the highest quality in all areas of pure and applied mathematics" (see www.ams.org/jams/aboutjams.html). Our sample is made up of Vol. 9 Number 4 (1996) and Vol. 15 Number 4 (2002). Its size is 0.29m words.

As can be seen, the bulk of our data dates from the 1990s, which we treated as a monolithic block.² An attempt was made to balance sample sizes across the journals, and to represent each year in the 1990s evenly. For EJ, the latter proved difficult due to document format problems, which is why for this journal only the years 1998–2000 were sampled. Notice that the samples we took from the 1965 and 1980 volumes are comparatively smaller than our 1990s samples. A somewhat pragmatic reason for this is that it is rather laborious to obtain digitized text data of material published prior to 1990.

Also note that we included *all* textual material in each journal issue, including book reviews, editorial notes, research articles, etc.³ In all, with over 9 million words, our corpus can claim to be quite representative of the scholarly writing subject to analysis in the present study; even our pre-1990 data subcorpus is quite sizable (c. 0.94m words). For purposes of comparison, two previous studies that are methodically similar to the present study (Chevalier and Hudson,

² This is because in the period from 1990 to 2000, no substantially or statistically significant changes in economic academic prose were observable.

³ While different text types – for instance, research articles and book reviews – differ in their form, they do certainly not differ much in their orientation as regards content and/or ideology, given that they are published in the same journal.

, Conrad, 1996) are based on much smaller databases: Chevalier and Hudson's corpus had a size of c. 0.25m words, Conrad's analysis is based on a corpus of 48,000 words.

Method

The present study will utilize corpus-linguistic methods to explore large-scale patterns of variance in economic and other academic journals. What are the characteristics of corpus-based analysis? According to a standard textbook,

the essential characteristics of corpus-based analysis are:

- it is empirical, analyzing the actual patterns of use in natural texts;
- it utilizes a large and principled collection of natural texts, known as a "corpus", as the basis for analysis;
- it makes extensive use of computers for analysis, using both automatic and interactive techniques;
- it depends on both quantitative and qualitative analytical techniques.

(Biber et al., 1998: 4)

In a nutshell, corpus-based studies are empirical analyses of how language is used, relying empirically on authentic, systematically sampled texts.⁴ The present study will rely on a standard method in quantitative corpus linguistics, namely, aggregating usage data of a portfolio of variables across the samples in our corpus, and analyzing the range of variance. This variationist discourse-analytical approach is methodically similar to multidimensional analyses of register variation, an approach developed by Douglas Biber and his co-workers (see Biber, 1988 for an overview).

All variables will be investigated by measuring their *text frequency*, that is, by measuring how often a given token turns up in the texts relative to text size. Our overall argument will rest on two key assumptions: (1) The samples we use are representative of the language used in the respective journals, and, by inference, of the language used in their discipline and/or their orientation (due to our corpus size, this assumption is unproblematic); and (2), the more often a token or word is used – and thus, the higher its text frequency is – the more prominently does the topic which the token denotes figure in academic discourse. Although (2) is a quite safe assumption and adopts linguistic standard practice in analyses of large amounts of data (see, for instance, Rayson et al., 1997), the reliance on vocabulary frequency has limitations: the method is blind, of course, to *how* a token is used and particularly to the stance an author may take towards that token. Yet, it is safe to assume that there is a positive correlation between the frequency with which an item is employed in a given journal and the degree to which the item is crucial for the scholarly community of which the journal is part of.

As for the envelope of variables subject to analysis, we will probe academic prose from both a lexical and stylistic perspective. This is another way of saying that we will investigate into the frequency of usage of a host of specific terminology (listed in Table 1 below), and, at the same time, that we will analyze quantitative variation across a number of stylistic features, that is, tokens that do not necessarily carry semantic meaning (see Table 2 below).

[insert Table 1 about here]

⁴ See Biber et al. (1998), Meyer (2002), or McEnery (2006) for readable introductions to the method.

The tokens in Table 1 are what we choose to call *topic markers*: they reflect *what* is being written about.⁵ These markers fall into four major groups:

1. *abstractive-mathematical topic markers*, which we take to be indicative of logical-hypothetical reasoning that proceeds through universal deduction and which is unrelated to observation or experience;
2. *socio-scientific topic markers*, which we consider to be typical of theory-driven, and often normative, discourse whose primary subject is human interaction;
3. *empirical topic markers*, which are indicative of strongly data-driven and inductive discourse;
4. *archetypally economic markers*, which we regard as signifying discourse which is characteristic of economics, using signal terms such as *rationality* and *equilibrium*.

[insert Table 2 about here]

Items in table 2, by contrast, are *style features*: they relate to *how* authors write about their subjects. We included the usual suspects known from previous linguistic research of stylistic variation (cf., for instance, Biber, 1988): probability markers and markers of authorial comment (POSSIBLE / CERTAIN / SURPRISING) and the modals OUGHT and SHOULD⁶, causal conjunctions as markers of explanative style, conditional conjunctions as markers of theorizing and hypothesizing style, and the first person pronouns I and WE⁷ as markers of personal style. While topic markers in Table 1, therefore, are key indicators of 'aboutness' and of terminology used, the style features in Table 2 are indicators of 'howness'.

In all, then, the database to be analyzed in the present study is a 13×24 matrix: it spans 13 measuring points (taken from representative samples of seven academic journals), with each measuring point being characterized by text frequency figures for 19 topic markers and five style markers.

Results

Analyzing the same corpus, Goldschmidt and Szmrecsanyi (to appear) offer an in-depth analysis of instructive differences between the journals sampled in terms of individual (groups of) topic and style markers. The present study seeks to complement this analysis by adopting a bird's eye perspective – indeed, a satellite perspective – on large-scale patterns and longitudinal linguistic shifts in economic journals. In this spirit, we will rely on two advanced statistical techniques to analytically engage our database, cluster analysis and multidimensional scaling.

⁵ Here and in the following, entries will have to be understood as linguistic lemmas; for instance, the lemma ASSUME subsumes all the formal lexical variations which may apply, such as *assuming*, *assumes*, or *assumed*.

⁶ Note that the semantics of these modals is actually many times more complex than is presented here. This caveat notwithstanding, probability and authorial comment (also called epistemic semantic loading) appears to be their core meaning.

⁷ As far as *we* is concerned, we are interested in the so-called exclusive *we* as opposed to the so-called inclusive *we*, which is why we manually corrected figures for inclusive *we*. In case of the former, authors refer exclusively to themselves, while in the case of the latter, they refer to themselves and their readership (Example for inclusive *we*: “Combined with the $j = 1$ case of Lemma 3.1, we obtain Z.”)

An aggregate view of large-scale patterns

To begin with, let us explore large-scale clusterings and patterns in our database, analyzing the variance in topic and style markers in one go. To obtain a snapshot of the big picture, we conduct CLUSTER ANALYSIS, a set of statistical techniques to objectively group a large number of cases (in this study, samples of academic prose) into a smaller number of discrete and meaningful clusters on the basis of some sort of similarity – in our case, similarities between text frequencies of topic and style markers. The method is widely used in disciplines such as biology, the social sciences (e.g. social network analysis), and, in the realm of linguistics, in dialectology (cf., for instance, Shackleton, 2005), dialectometry (for example, Goebel, 2004), and in corpus linguistics (cf. Gries, to appear); for an introduction to the technique from the social scientist's perspective, see Aldenderfer and Blashfield (1984).⁸ Data clustering can be visualized by tree diagrams, also known as DENDROGRAMS, where one finds individual samples on the left and successively larger clusters as one moves rightwards; the cluster analysis run on our database yields the dendrogram in Figure 1. Essentially, dendrograms work in much the same way as family trees: in a dendrogram, much as in a family tree, the closer to the left we find a split between a pair of samples, the more closely related linguistically the samples are. In what follows, we will embark on a step-by-step exploration of the dendrogram in Figure 1, working our way from the right to the left.

[insert Figure 1 about here]

As can be seen, the most fundamental split occurs between the JAMS and the rest of the journals sampled. This means that the formal-abstractive, mathematical discourse sampled in the JAMS is a quite unique genre that is fairly distinct, in particular, from the genre that is exhibited in the economic journals in our sample. Next, the database is split up into three clusters (indicated by the dotted line and the shaded branches in Figure 1) such that the JAMS remains a distinct cluster (cluster 3) while the other samples in the database are partitioned into two groups:

- Cluster 1 comprises the three AJES samples, the contemporary sample of the JoEP, the 1965 sample of the AER, and the JOM (but note that the economic journals are, predictably, more similar to each other than to the JOM). In essence, therefore, the economic journals in this cluster share an affinity to the JOM, which might indicate that these specimens of economic prose are rather empirically oriented, descriptive, and application-oriented in nature. At the same time, notice that the three longitudinal samples of the AJES form a fairly homogeneous group.
- Cluster 2 is less homogeneous than cluster 1 and spans the more recent samples of the AER, the EJ, and, crucially, the AJS. Thus, cluster 2 appears to be a rather socio-scientific cluster whose component samples might be interpreted as being somewhat theoretical in nature, analytical, and method-oriented. Also note that the cluster can be further subdivided into two sub-clusters, one comprising the 1980 sample of AER, the 1990 sample of EJ, and the AJS; and the other one containing the earlier EJ samples (1965, 1980) and the contemporary AER sample.

⁸ On a technical note, we utilized a hierarchical agglomerative clustering algorithm to partition the samples in our database into clusters, using between-groups-linkage as clustering method and square Euclidean distance as interval measure. This means that the clustering algorithm started out with individual samples, merging them stepwise – when possible – with other samples to arrive at successively larger clusters under the condition that cluster members be maximally similar.

In all, then, we have seen in this section that in the aggregate view – considering joint variance of topic *and* style markers – economic journals are considerably closer to socio-scientific and applied natural science genres than to the mathematical, formal-abstractive genre sampled in the JAMS. Among the samples taken from economic journals, some of these pattern rather with the socio-scientific AJS, others with the applied, natural-scientific JOM. We shall engage in a more thoroughgoing investigation of these affinities in the next section.

Similarities and dissimilarities in a longitudinal perspective

We will now turn our attention to a more fine-grained picture of the shifts and changes that have affected economic academic prose in the course of the past 40 years, viewed against the benchmark of contemporary issues of the AJS, the JOM, and the JAMS. Accordingly, in this section we investigate topic markers and style markers separately utilizing MULTIDIMENSIONAL SCALING (henceforth: MDS) to visualize similarities and dissimilarities between the samples in our corpus. MDS is a set of statistical techniques, now popular in empirical linguistics (cf., for instance, Nerbonne et al., 1999), designed to uncover the hidden structure of variation in a dataset like ours (for an accessible introduction to the technique, see Kruskal and Wish, 1978). In terms of the present study, this means that we are going to scale down to two dimensions the 19 original topical dimensions and five style dimensions by which every journal sample in our database is characterized⁹, visualizing aggregate (dis-)similarities as well as longitudinal developments in two-dimensional maps. The obvious advantage here is that because such visualizations use the straightforward concepts of space and distance, they can be interpreted fairly intuitively. Much as with geographic maps, the further two points are apart, the more dissimilar (in geographic terms, distant) they are. If two pairs of points are equidistant, the pairs of journal samples they represent are equally (dis-)similar. As a matter of fact, one could apply a ruler to a MDS map and gauge relative similarities by simply measuring spatial distances.

We start by discussing longitudinal changes in the use of topic markers; Figure 2 displays the corresponding MDS visualization. First and foremost, it is striking that economic journals, now and in the past, quite saliently form a topical genre of their own that is distinct from each of the non-economic journals in our corpus. Still, the visualization in Figure 2 squares with the dendrogram in Figure 1 in that the economic journals are considerably closer to the socio-scientific genre in AJS and to the descriptive natural science genre in the JOM than to formal-abstractive, mathematical genre in the JAMS. This is converging evidence that the JAMS can be considered an outsider genre in our corpus.

[insert Figure 2 about here]

Turning to longitudinal shifts, the economic journals in Figure 2 clearly exhibit, on aggregate, a rather noticeable drift to the right of the diagram (indicated by the dotted arrows) – that is, a shift away from the topical coordinates of the JAMS and the JOM and towards the topical coordinates of the AJS. Still, there are instructive differences between the economic journals in terms of the exact nature of this drift: while the AJES has remained most beholden to the genre represented by our JOM sample, the EJ is now most similar, among the economic journals sampled, to contemporary sociological academic discourse as evidenced in the AJES. The AER, on the other hand, has traveled the farthest distance overall – maybe because the journal aims to be particularly responsive to new trends and fashions in economics – and

⁹ Technically speaking, we derived perceptual maps from Euclidian distances (cf. Young and Harris, 1992).

especially so in the period between 1965 and 1980. Observe, along these lines, that while the AER covered the middle ground, in a way, between the AJES and the EJ in 1965, its 1990s sample occupies an extreme position to the right of Figure 2. What is more, the path along which the AER has traveled topically is not necessarily a trajectory away from the JAMS and towards the AJS, but rather one (somewhat orthogonal to the JAMS–AJS axis) that maximizes distance from the descriptive natural science discourse in the JOM. Last but not least, the economic journal that in the 1990s appears to cover the economic middle ground in the topical triangle spanned by the non-economic journals is demonstrably the JoEP.

[insert Figure 3 about here]

Does our corpus testify to significant longitudinal shifts with regard to style markers as well? The MDS visualization in Figure 3 suggests it does. Quite arrestingly, some of the shifts in style marker usage actually contradict the developments we had observed with regard to topic markers. This is another way of saying that in terms of style markers, there is *not* a general drift of economic journals towards the socio-scientific style of the AJS. Visual inspection of Figure 3 moreover reveals that in regard to style, academic journals do not really appear to form an academic genre of their own (recall that they do in terms of topic). Let us now discuss in detail the style changes observable for the three economic journals for which our corpus provides longitudinal data:

- Authors publishing in the AJES stylistically shifted away from the AJS, in the direction of the more formal-abstractive style of the JAMS, in the period between 1965 and 1980. This shift was more or less reversed, though, in the period between 1980 and the 1990s. As for contemporary economic texts, the AJES is now the economic journal with the highest stylistic affinity to the AJS. At the same time, notice that of the three economic journals sampled longitudinally, the AJES has been the one most stable stylistically over time.
- Material published in the EJ did not change significantly over time in terms of its relative stylistic distance, or closeness, to the AJS, but it is now considerably more similar stylistically to the mathematical-abstractive style exhibited in the JAMS and to the descriptive natural-scientific style displayed in the JOM. As a matter of fact, the contemporary material in EJ is rather close to the style of our 1980 sample of the AER.
- In lockstep with variance in topic markers, the AER has experienced the most dramatic shifts in its authors' stylistic preferences over time. While the journal's relative similarity, or dissimilarity, to the socio-scientific style of the AJS has not altered substantially, the AER has become stylistically a great deal more similar to mathematical academic prose (as represented in the JAMS) in the period between 1965 and 1980. Somewhat similar to the AJES, this drift has been partially reversed in the period between 1980 and the 1990s. Moreover, it seems worth pointing out that while our 1965 sample of the AER was exceedingly similar to the descriptive natural-scientific style of the JOM, our 1990s sample of the AER is the one economic sample in our corpus most dissimilar to the JOM's style. This movement of AER's away from the stylistic preferences exhibited in the JOM squares very neatly with the topical divergence between the AER and the JOM discussed above. It is also interesting that according to our data, textual material in the EJ and in the AER has been subject to stylistic drifts in more or less diametrically opposite directions over time.

In all, our analysis in this section would seem to have offered that economics is a rather agile academic genre. While the topics – or, at a minimum, the terminologies – subject to discussion in economic journals have drifted in the past 40 years towards the coordinates of

contemporary material in the AJS and away from the topics (or terminologies) recurrent in the JAMS, the stylistic preferences of authors publishing in economic journals have overall come to display an increased affinity to the mathematical journal in our corpus, the JAMS.

Conclusion

On the first page of *The Rhetoric of Economics*, McCloskey draws attention to the following observation:

The economic conservation has heard much eloquent talk, but its most eloquent passages have been mathematical. Especially since the 1930s economists of all schools have become enchanted by the new and scientific way of talking. Most journals of economics nowadays look like journals of applied mathematics or theoretical statistics. (McCloskey 1985, 3)

Our analysis has suggested that McCloskey is, indeed, correct – at least as far as economists' *stylistic preferences* are concerned. Given our data, it seems to us that economists aspire to be (rhetorical) adherents of a mathematical science. But in terms of substance, this is, alas, wishful thinking: the issues and questions – the *topics* – that concern economists cannot be conclusively dealt with using mathematical approaches alone, as our findings have indicated. Economists, in short, have certainly come to talk the talk, but they cannot, in fact, walk the walk. On the contrary: economists might have to let go of the mathematically inspired 'just so stories' of orthodox economics. Instead, it is likely that we have to turn back to the tedious business of economic methodology again, even though it is clear that we cannot abandon our mathematical hump: "And from that day to this the Camel always wears a hump (we call it 'hump' now, not to hurt his feelings); but he has never yet caught up with the three days that he missed at the beginning of the world, and he has never yet learned how to behave" (Kipling, 2002 [1902]: 16).

References

- Aldenderfer, Mark S., and Blashfield, Roger K. 1984. *Cluster Analysis: Quantitative Applications in the Social Sciences*. Newbury Park, London, New Delhi: Sage Publications.
- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, Conrad, Susan, and Reppen, Randi. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Chevalier, Gisèle, and Hudson, Richard. 2001. The Use of Intentional Language in Scientific Articles in Finance. *Journal of Economic Methodology* 8:203-228.
- Conrad, Susan. 1996. Investigating academic texts with corpus-based techniques. *Linguistics and Education* 8:299-326.
- Goebel, Hans. 2004. Sprache, Sprecher, und Raum: Eine kurze Darstellung der Dialektometrie. Das Fallbeispiel Frankreich. *Mitteilungen der Österreichischen Geographischen Gesellschaft* 146:247-286.
- Goldschmidt, Nils, and Szmrecsanyi, Benedikt. to appear. What do economists talk about? A linguistic analysis of published writing in economic journals. *American Journal of Economics and Sociology*.
- Gries, Stefan Th. to appear. Exploring variability within and between corpora: some methodological considerations. *Corpora*.
- Kipling, Rudyard. 2002 [1902]. *Just So Stories*. New York: Signet Classics.
- Kruskal, Joseph B., and Wish, Myron. 1978. *Multidimensional Scaling: Quantitative Applications in the Social Sciences*. Newbury Park, London, New Delhi: Sage Publications.
- McCloskey, Deirdre. 1985. *The rhetoric of economics*. Madison, WI: University of Wisconsin Press.
- McCloskey, Donald. 1991. Storytelling in Economics. In *Economics and Hermeneutics*, ed. Don Lavoie. London: Routledge.
- McEnery, Tony, Xiao, Richard, and Tono, Yukio. 2006. *Corpus-Based Language Studies: An Advanced Resource Book*. London, New York: Routledge.
- Meyer, Charles F. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Nerbonne, John, Heeringa, Wilbert, and Kleiweg, Peter. 1999. Edit Distance and Dialect Proximity. In *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, eds. David Sankoff and Joseph B. Kruskal. Stanford: CSLI.
- Rayson, Paul, Leech, Geoffrey, and Hodges, Mary. 1997. Social Differentiation in the Use of English Vocabulary: Some Analyses of the Conversational Component of the British National Corpus. *International Journal of Corpus Linguistics* 2:133-152.
- Schroeder, Jeanne L. 2001. Just So Stories: Posnerian Methodology. *Cardozo Law Journal* 21:351-423.
- Shackleton, Robert G. Jr. 2005. English-American Speech Relationships: A Quantitative Approach. *Journal of English Linguistics* 33:99-160.

<i>discourse type</i>	<i>item</i>	<i>example</i>
abstractive- mathematical	ASSUME/ ASSUMPTION	"We <i>assume</i> risk neutrality and independence of relevant factors." (AER Vol. 80, 1990)
	PROOF	"[...] no <i>proof</i> has been presented that this hypothesis is the correct one [...]" (AJES Vol. 80, 1965)
	THEOREM	"[...] the above result is stated as a lemma instead of as a <i>theorem</i> ." (AER Vol. 89, 1999)
socio-scientific	CULTURE/ CULTURAL	"Byelorussians were and remain <i>culturally</i> Russian" (AJES Vol. 60, 2001)
	HISTORY/ HISTORICAL	"This period in the <i>history</i> of economic thought stands out prominently [...]" (AJES Vol. 50, 1991)
	INSTITUTION/ INSTITUTIONAL	"Neither did he outline a theory of governance with <i>institutional</i> procedures [...]" (AJES Vol. 60, 2001)
	JUSTICE	" <i>Justice</i> may even drop out of sight when urgency presses upon the government." (AJES Vol. 50, 1991)
	SOCIETY	"[...] to protect the kind of <i>society</i> we believe to be worth saving for future generations [...]" (AJES Vol. 50, 1991)
	THEORY/ THEORETICAL	"[...]a good economic <i>theory</i> would identify a specific cause for a societal problem [...]" (AJES Vol. 50, 1991)
	METHODOLOGY/ METHODOLOGICAL	" <i>Methodology</i> is [...] the study of the grounds for employing methods." (AJES Vol. 59, 2000)
empirical	DATA	"Unfortunately, however, there are no <i>data</i> on within-country trade." (AER Vol. 89, 1999)
	EVIDENCE	"Is there any <i>evidence</i> that economists behave as if they were maximizing equation (I)?" (AER Vol. 70, 1980)
	METHOD	"The [...] procedure used, most commonly the simplex <i>method</i> , normally is completed by a computer." (AJES Vol. 50, 1991)
	OBSERVATION/ OBSERVE	"[...] exogenous technological developments might have amplified some <i>observed</i> changes." (AER Vol. 89, 1999)
	EFFECT	"Let us turn, then, to MM's analysis of the <i>effect</i> of taxes." (AER Vol. 55, 1965)
	MODEL	"This is proved in two different <i>MODELS</i> of uncertainty." (AER Vol. 70, 1980)
	RESULT	"This <i>result</i> also holds under certainty." (AER Vol. 70, 1980)
archetypally economic	EQUILIBRIUM	"Our strategy is to specify an <i>equilibrium</i> MODEL of the auto industry." (AER Vol. 89, 1999)
	RATIONALITY/ RATIONAL	"The neoclassical theory then assumes only one type of [...] homo oeconomicus, who is <i>rational</i> ." (AJES Vol. 60, 2001)

Table 1: Topic markers studied

<i>function</i>	<i>item</i>	<i>example and hypotheses</i>
probability markers, authorial comment (see Swales 1990: 136)	expressions of authorial evaluation and/or intrusion (POSSIBLE/ CERTAIN/ SURPRISING)	"His unfortunate choice of terminology <i>possibly</i> served to alienate a considerable body of opinion" (AJES Vol. 60, 2001)
	modals of expectation, advice, and probability (OUGHT/ SHOULD)	" <i>Should</i> 'ownership' and 'property' be so defined as to refer to a single essential right?" (AJES Vol. 60, 1965)
explanative style suggesting high intellectual complexity (see Swales 1990: 115)	causal conjunctions (<i>because</i>)	"Hence mathematical determinations are not judgments, <i>because</i> mathematics is an a priori discipline." (AJES Vol. 50, 1991)
theorizing and hypothesizing style	conditional conjunctions (<i>if/ provided that/ in case</i>)	" <i>If</i> not sterilized, this will lead to a credit crunch [...]." (AER Vol. 89, 1999)
personal style (see Swales 1990)	first person pronouns (<i>I/ we</i>)	"[...] <i>we</i> feel that our assumption represents a good first approximation." (AER Vol 55, 1965)

Table 2: Style markers studied

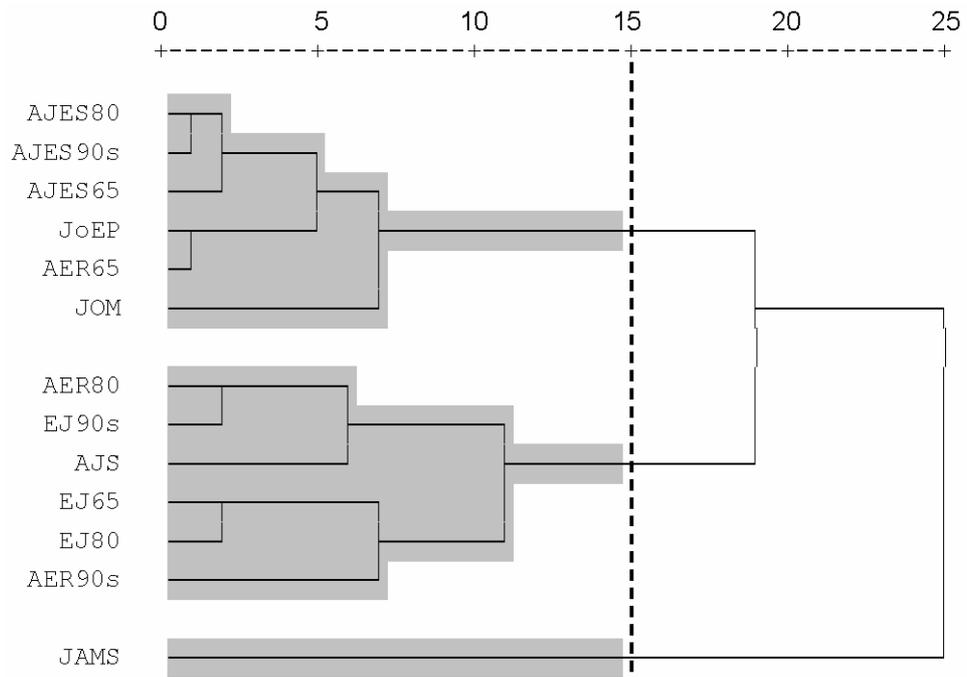


Figure 1: Dendrogram derived from hierarchical agglomerative cluster analysis of 13 samples of academic prose, each jointly characterized by 19 topic markers and 5 style markers

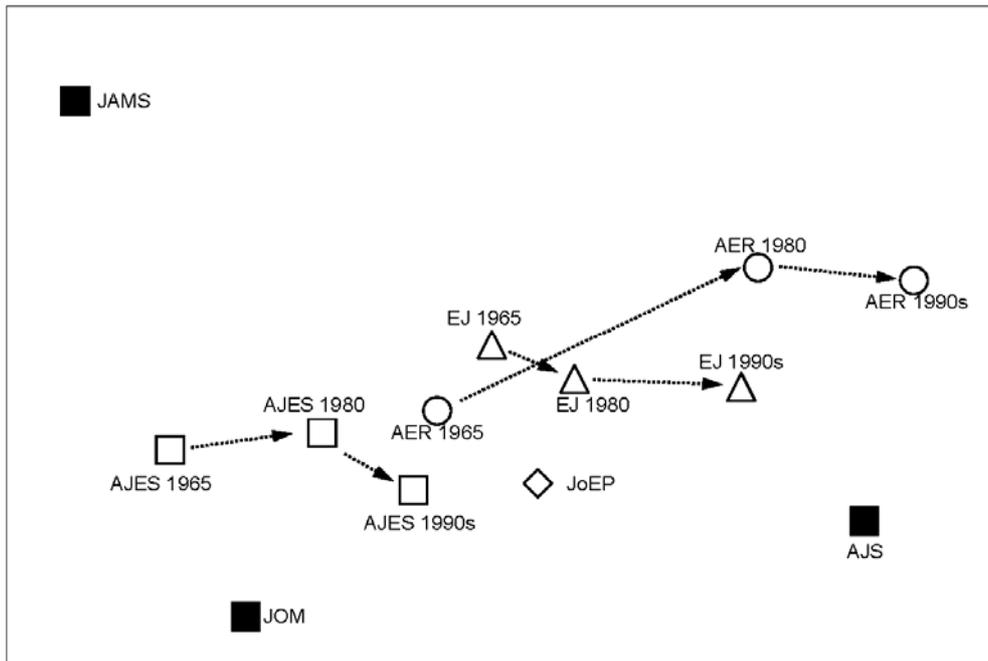


Figure 2: Topic markers – MDS visualization of 13 samples of academic prose, each characterized by a vector of 19 topic markers. Dotted arrows indicate longitudinal shifts

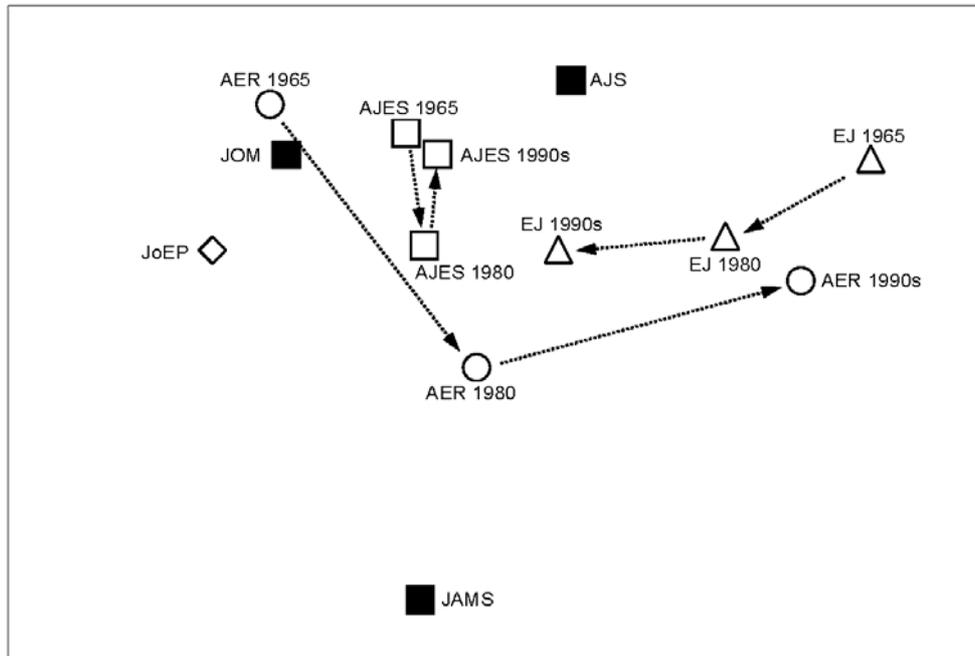


Figure 3: Style markers – MDS visualization of 13 samples of academic prose, each characterized by a vector of 5 topic markers. Dotted arrows indicate longitudinal shifts