

Jacques Lacan's Registers of the Psychoanalytic Field, Applied using Geometric Data Analysis to Edgar Allan Poe's "The Purloined Letter"

Fionn Murtagh¹ and Giuseppe Iurato²

¹University of Derby, UK; and Goldsmiths University of London, UK

²University of Palermo, Italy

Corresponding author: Fionn Murtagh , fmurtagh@acm.org

Abstract

In a first investigation, a Lacan-motivated template of the Poe story is fitted to the data. A segmentation of the storyline is used in order to map out the diachrony. Based on this, it will be shown how synchronous aspects, potentially related to Lacanian registers, can be sought. This demonstrates the effectiveness of an approach based on a model template of the storyline narrative. In a second and more comprehensive investigation, we develop an approach for revealing, that is, uncovering, Lacanian register relationships. Objectives of this work include the wide and general application of our methodology. This methodology is strongly based on the "letting the data speak" Correspondence Analysis analytics platform of Jean-Paul Benzécri, that is also the geometric data analysis, both qualitative and quantitative analytics, developed by Pierre Bourdieu.

Keywords

Text mining, narrative, Lacan registers, real, imaginary, symbolic, Correspondence Analysis

I INTRODUCTION

1.1 Psychoanalytical Use of Edgar Allan Poe Story

Edgar Allan Poe's "The Purloined Letter" is a story that is investigative and elaborative. It is not just explanatory, reducing the case study in this story to facts and assertions that are ordered. Rather, it is also elucidatory, and positioning in a larger, broader, contextual picture. Also this allows to identify better where truth lies by means of simple psychoanalytic tools. Positioning is done through contextualization. We hypothesize that any and all such elucidation, and contextual positioning, is potentially relevant for various domains such as theatre and drama, legends and mythology, and *mutatis mutandis*, for poetry and music.

Therefore, Poe's story is not simply the investigation of illegal behaviour. There are parallels and analogies drawn with schoolboys playing with marbles and strangely enough with mathematical reasoning. These strange connections are just possible in the unconscious realm. A lot of foremost thinkers have discovered, or at least viewed, very interesting mappings of Poe's story into the most interesting contexts. See Poe [2016] for discussion of Michel Foucault, Jacques Lacan, Jacques Derrida and others.

Description follows of the psychoanalytical approach developed by Lacan, encompassing analysis of synchrony and of diachrony. Diachrony can be based on the inducing of a segmentation of the narrative or storyline into a sequence of main scenes or acts. The synchronous elements

decompose any act by means of the three Lacanian registers or orders of the so-called psychoanalytic field in which every human event performs at the unconscious level. The three Lacanian registers, comprising the *psychoanalytic field*, are the real, the imaginary and the symbolic. Lacanian psychoanalysis seeks to outline the co-participation of these three registers in each event and subject of the story, but with a synchronic predominance of one over the others, which will then be the one that is diachronically identifiable.

This study has the following objective. Firstly we seek to reveal or to determine Lacan's registers in a highly realistic case study.

Our mapping of Lacan's registers in the Poe story leads to Section VI. Specifically seeking metaphor and metonymy is at issue in sections VII to IX.

1.2 Source of Data and Preparation

In this subsection, and throughout this paper, we detail the data processing carried out, firstly for reproducibility of this study, and secondly for all aspects relating to generalization of this work, and application to other textually expressed content.

The Edgar Allan Poe text of "The Purloined Letter" was taken from Poe. Accented characters required correction, following the 1845 editions in Poe [1845], Poe [1845 (MDCCCXLV)]. A program was run on this text that determined sentence boundary (using a full stop), and also took into account blank lines that indicated paragraph boundary. Some cases of repeated dashes, repeated dots, exclamation marks and question marks were modified manually in the input text. The processing allows the specification of standard contractions that are not to be taken as sentence boundaries. (The following were at issue in regard to being ended with a full stop or period but this did not connote the end of a sentence: *no*, *No*, *C*, *G*, *St.*) A CSV (comma separated values) formatted file was created, with the sentence sequence number, the paragraph sequence number, and the sentence content. This led to 322 sentences and 123 paragraphs. For each paragraph, the speaker was also noted: the Narrator, Dupin and the Prefect. In section 1.3, some further background description on the Poe story will be provided.

1.3 Dramatis Personae

The characters in this short story are as follows: (1) C. Auguste Dupin (young private detective); (2) Monsieur "G – –", or G. or Prefect (police chief); (3) the narrator (Dupin's friend and roommate); (4) the Minister "D – –", or "the D – –", or the minister (the villain); (5 and 6) the personage [in the royal boudoir], or other unnamed royal person (often considered as Queen, King); and (7) "S –", sender of the letter (only one occurrence of this name).

Examples follow of the first and the last sentences.

- First: "At Paris, just after dark one gusty evening in the autumn of 18 – , I was enjoying the twofold luxury of meditation and a meerschaum, in company with my friend, C. Auguste Dupin, in his little back library, or book-closet, au troisième, No. 33 Rue Dunôt, Faubourg St. Germain."
- Last: "They are to be found in Crébillon's 'Atrée'"

1.4 Brief Background on the Geometric Data Analysis Methodology

Our approach is influenced by how the leading social scientist, Pierre Bourdieu, used the most effective inductive analytics developed by Jean-Paul Benzécri. See Le Roux and Rouanet [2004], Grenfell and Lebaron [2014], Lebaron and Roux [2015]. This family of geometric

data analysis methodologies, centrally based on Correspondence Analysis encompassing hierarchical clustering, and statistical modelling, not only organises the analysis methodology and domain of application, but even integrates them. The second in a set of principles for data analytics, listed in Benzécri [1973] (page 6), included the following: “The model should follow the data, and not the reverse. ... What we need is a rigorous method that extracts structures from data.” Closely coupled to this is that (Benzécri [1983]) “data synthesis” could be considered as equally if not more important relative to “data analysis”. Analysis and synthesis of data and information obviously go hand in hand.

The work of Andreas Schmitz, dealing with Angst and fear (Schmitz [2015], Schmitz and Bayer [2014]), links together Freud and Bourdieu, for example, in regard to “libido within habitus-field theory”. Among the conclusions in Schmitz [2015] are how we have:

1. “Libido constitutive for the foundational concepts of *habitus* and *fields*”.
2. Janus-faced character of libido: interest and Angst as constitutive moments of (i) Habitus and practice, (ii) Social space and social fields, and (iii) Symbolic domination,

In Schmitz and Bayer [2014], also presented in Schmitz [2015], Schmitz notes the limits of statistical linear modelling for relating personality factors in social space. Moving beyond that methodology, there is categorical interest and personality types, accompanying the socio-structural information for the geometric construction of social space. The aim is to demonstrate in general, whether psychological characteristics will correspond with the structure of social space in a discontinuous way. (This summarizes perspectives in Schmitz and Bayer [2014], p. 11. The following is from p. 14.) Habitus defines the nexus between structure and subject, whereby the correspondence of social position and “psychic” disposition are understood as class-specific, and thus discontinuous. Psychiatric indicators are used in a discontinuous way (as befits such categorical variables). From a psychoanalytic viewpoint, the habitus roughly corresponds to Freudian super-ego agency, hence it belongs to the Lacanian symbolic register. So, studying the latter, we might infer features of habitus, hence answer to the above issue regarding links between psychological characteristics and social structure. We shall focus on the linguistic.

II FIRST STUDY: ANALYSIS USING SIMPLE DIACHRONIC MODEL

Below, in this paper, most of the set of words in the Poe text are used. This is so as to take account of emotion and sentiment, expressed language-wise through adjectives and adverbs, and so on. Also below, text-based, i.e. data-based, story or narrative flows are considered.

In this first study, a somewhat simplified diachronic model of the Poe story is used. That is, a model of the evolution or flow of the story is used. This is strongly based on a Lacanian interpretation. Also in this first study, from the text of the Poe story, nouns are used. This is in order to have a relatively quick, first view of the relationship between key terms.

We consider now, the Lacanian motivation, and indeed justification, for this work.

Lacan’s *psychoanalytic field* is structured into three dimensions or orders, termed the *Lacan registers*, which may be considered as components of this field, closely linked to each other (*Borromean knot*). These are the *symbolic register*, the *imaginary register* and the *real order*. The Lacan psychoanalytic field relies on unconscious realm.

The *symbolic register* is that field component in which *signifiers* act, operate and combine according to laws and rules of structural linguistics, above all the negation. The main law of this

Lacan registers	Persona
Act 1	
Real	King unaware. Inconceivable content of the letter.
Imaginary	Queen worried about letter and its content. This was then hidden.
Symbolic	Minister seizes letter using apparent substitution with own letter.
Act 2	
Real	Police also unsuccessful. Police were on Queen's request. Hence unaware. [Required solution: link between real and symbolic]
Symbolic	Dupin, having his aim disguised, sees probable letter; returns; seizes letter using apparent substitution with own letter.
Act 3	
Real	Minister unaware, could be threatened also by this affair.
Imaginary	Dupin replacement letter had sinister sentence.
Symbolic	Here: the letter, the signifier, in its circuit. [It was/is real; the imaginary was associated with it; symbolic related to apparently similar letters, and also being related to various associated contexts.]

Table 1: Very summarized rendition of the Poe story. Summary of participant roles, relative to Lacanian registers.

field component is the so-called *Name-of-the-Father*, which triggers the formation of the *signifiers chain*. This register is the most prominent one in acting on the individual, through the intervention of imaginary register.

The *imaginary register* is that field component which springs out of the unconscious apprehension of ones own bodily image of the child (*mirror stage*) on the basis of the primary dual relationship of identification with ones own mother. It is the basis for the growth, by *alterity*, of the Ego agency and the narcissistic pushes, when mother, through Name-of-the-Father law, casts the child into the symbolic register, naming her or him.

The *real order* is that field component which is defined only in relation to symbolic and imaginary registers, where there is all that impossible, unbearable or inexpressible content expelled or rejected by these latter two registers. The symbolic and imaginary registers, together the real order, are in relationship to each other, mostly in opposition.

In Table 1, there is a useful, very summarized, rendition of the Edgar Allan Poe story. It is structured as what we label here as the succession of Acts 1, 2, 3. One register will dominate others synchronically, i.e. at any given time-point. The symbolic register will win out, in that there is a fairly natural progression from the real, to the imaginary, thereby resulting in the symbolic. The real register is occupied by what the symbolic ejects from reality, and that cannot be formalized by language.

In this first study, the Poe story consists of 321 sentences, and a corpus of 1741 words. These words are of length at least 1, all punctuation has been removed, and upper case has been set to lower case. Then we require a word to be present at least 5 times, and used in at least 5 sentences. Next, words in a stopword list were removed. These are (definite, indefinite)

articles and common parts of verbs, and such words (using the *tm*, text mining, package in the R software package). Single letter words were also deleted (e.g. “s” resulting from “it’s”, or “d” resulting from “didn’t”, when the apostrophe here was replaced by a blank). Then just nouns alone were selected. There were 48 nouns at this stage. Some of the 321 sentences became empty. There were 213 non-empty sentences, as noted, crossed by 48 nouns. In the 213 sentences, there were 424 occurrences of these words.

The sentence set, characterized by words used, endowed with the chi squared metric is mapped, using Correspondence Analysis, into a Euclidean metric-endowed factor space. In order just to retain the most salient information from this semantic, factor space, we use the topmost 5 axes or factors. These 5 axes account for 17.75% of the inertia of the sentences cloud, or identically of the nouns cloud. Figure 1 displays the hierarchical clustering of the sentences, that are in their 5-dimensional semantic or factor space embedding. The complete link agglomerative clustering criterion permits adherence to the sequential order of the sentences (Murtagh et al. [2009], Bécue-Bertaut et al. [2014], Legendre and Legendre [2012]).

To follow our template of three acts, we take the three largest clusters. In the dendrogram in Figure 1 we therefore have the partition, containing three clusters, close to the root node. These clusters relate to sentences 1 to 53, sentences 54 to 151, and sentences 152 to 213. These are to be now our acts 1, 2, 3, following the template set out descriptively in Table 1. The number of sentences in each of these acts is, respectively, 53, 98, 62. Next for analysis, we create a table crossing 3 acts by the noun set of 48 nouns.

The complete factor space mapping is just in 2 dimensions. Figure 2 displays the words that have the highest contribution to the inertia of this plane. To see the relationship between the words that are close to the origin, thus essential to the whole of the narrative line, to all acts, Figure 3 displays the region of the plane that is close to the origin. We see “letter” and other words.

We can try to investigate the internal structure of our “template” acts. Figure 4 displays the hierarchy (using the appropriate agglomerative criterion of Ward’s minimum variance) constructed in the 5-axis or 5-factor embedding of this data. From left to right here, the three clusters resulting from the dendrogram cutting, or slicing into a partition, as displayed, correspond mapping-wise to act 2, act 3, act 1. Cf. what is displayed in Figure 2.

With the perspective of Lacan’s registers we could look at a set of three clusters in each of these acts. Let us look at the leftmost cluster here. We can read off the following three clusters: first cluster, “reward, boy, case, school, furniture, microscope”; second cluster, “secret, course, thing”; third cluster, “chair, individual, book”. This has just been reading off three fairly clearly determined clusters. Of course we can see that the second cluster and the third cluster are merged fairly early on in this agglomerations.

Rather than attempting to relate these clusters with Lacan’s registered, let us instead just draw the following conclusion. In this first study, it has been shown how a template of segmentation can be easily considered. So the diachrony can be investigated. In our opinion, the retained words consisting of nouns are a good way to focus our study, and also while we succeeded well in imposing our template of the segmentation of the Poe story into three acts. However while they certainly lead to interesting perspectives, for general-purpose use of this methodology, it would be preferable to allow for a somewhat more open perspective on the data. This we do next, analysis of diachrony and of synchrony, both newly investigated.

III LACANIAN REGISTERS EXPRESSED IN THE NARRATIVE FLOW

3.1 Lacanian Framework

Fundamental aspects of Lacanian methodology encompass the following.

- Metonymy, e.g. the name of the cause is used for denoting the effect or the object; it is associated with, and expressed by, diachronicity. Diachronicity horizontally combines patterns into metonymy. Finally, metonymy is to be associated with (Freudian) unconscious displacement, shifting and moving, under the pushes or drives of desire, the various signifiers, without an end but rather aimed always at seeking the lost object (lacking for the human) which escapes every signification.

Metaphor, e.g. a word or term is replaced with another similar or analogous term; through selection, metaphor is enabled by synchronicity. Synchronicity vertically selects patterns into metaphor. Finally, metaphor corresponds to (Freudian) unconscious condensation, which disguises and upsets meanings, until reaching the deepest unconscious levels. Metonymy and metaphor are the two main (Freudian) pathways of semiotic action.

- Signifiers will constitute the symbolic register. These signifiers combine like the basic structural elements of a language. The signifiers of the symbolic register undergo the rules of metonymy and metaphor.

For Lacan, the signifier dominates the signified, and not vice versa (as for De Saussure), through certain structural rules (similar to the linguistic ones) in which the former (signifiers) link together to give rise to signifier chains. Signifier chains are diachronic combination of signifiers synchronically selected, in which the signifiers follow each other oppositionally, like the words of a phrase.

Indeed, (synchronic) selection includes the case where a signifier excludes another one but remains in relationship with that other signifier, at least negatively, according to Aristotelian logic. These signifier chains will acquire then a conscious meaning following usual grammatical rules.

- The Imaginary is a register complementary to the Symbolic one. Generally, it is the realm of images and of the sensible representations (mostly, the visual ones) which mark our own lived experience. Imaginary fantasies and representations (thing representations) belong to the imaginary register as well, which will prepare the ground for the subsequent word representation.
- The Real is not reality as this is usually meant, that is to say, the world of everyday experience, which is already characterized by images and symbolic language, but it rather deals with the primary, rough experience of what is still not symbolized or imagined, with the impossible, that is to say, what is impossible to inscribe in every symbolic system or however represented in any possible imaging form.

3.2 Narrative Flows

All discourses, happenings, history, etc. are narratives, with one or more, and often many, narrative flows. In the narrative, there are various chronologies that may be investigated as sub-narratives. These include the sequence resulting from: (i) sections, (ii) speaker or agent, (iii) time or date or location, (iv) statistical segmentation into sections. The latter may be through syntax and style based clustering since tool words (function words) predominate. To the above can be added: (v) sentences, (vi) paragraphs. All this, is the result of the diachronic nature of the discourse, which, therefore, is explainable through Lacan theory. In particular, Lacan points out that, in the symbolic register, the diachronic selection axis of discourse is closely related with synchronic combination of signifiers which gives rise to the diachronic meaning, or signified, of

the discourse. These combination and selection processes, taking place in the symbolic register, are greatly influenced by the real register and, especially, by the imaginary register. These latter both push on the former.

We seek the most enlightening or the most illustrative of these narrative flows. By enlightening, we intend: seeking or determining specific outcomes. By illustrative, we intend: detecting or observing dialectical movement, or Aristotelian logic, or unconscious mind processes.

We are most interested in (i) metaphor, being an indicator of unconscious mind processes, for its synchronic nature, and (ii) metonymy, i.e. a term indicating diachronic employment (or use), that can be, therefore, transfer and handover.

Following the mapping of the text story into a semantic space, in regard to combinations of signifiers according to Lacan, for (i) metaphor, due to its synchronic nature, we use clustering. While, for (ii) metonymy, due to its diachronic nature, we use sequence constrained, i.e. chronologically constrained clustering. In (i) our aim is close association, expressed by highly compact clusters, while in (ii), we may consider varied chronological flows.

Our semantic analysis starting point is the set of all interrelations between narrative flow segments, belonging to the diachronic selection axis, and the words selected and retained, belonging to synchronic combination axis. We have that: “One terms the distribution of a word the set of its possible environments” (Benzécri [1982]).

IV TEXT ANALYSIS: INITIAL EXPLORATORY PROCESSING STAGES

The Poe story, in our text formatting, consists of 322 sentences, arranged as 123 paragraphs. As noted above, paragraph here is defined as text segments that are separated by blank lines. That includes vocal expressions, perhaps with some additional explanatory text, and also it may be noted that a few of the vocal expressions can be quite short. Nonetheless it is clearly the case that the paragraphs form useful text, and narrative, segments.

Next we also considered a segmentation into 8 sections, based on a reading of the Poe story. The introduction part of the story had 19 paragraphs. The initial outlining of the essential story, relating to the purloined letter, told by the Prefect with dialogue elements from the narrator and from Dupin, constituted section 2, with 26 paragraphs. Section 3, with 28 paragraphs recounted the Prefect’s search of the Minister’s hotel room. Section 4, with 14 paragraphs, takes place one month later, detailing the revelation that Dupin could provide the letter to the Prefect. Then section 5, with 6 paragraphs, starts off the background explanation by Dupin to the narrator. Section 6, with 16 paragraphs, continues in great detail as Dupin provides explanation to the narrator. Section 7, with 8 paragraphs, is the core of the storyline, where Dupin explains how he found the letter, how this was verified by him, and how he took hold of it in the following morning, putting what is referred to as a facsimile in its place. Finally section 8, with 6 paragraphs, is the explanation of, and justification for, the replacement of the letter by a facsimile.

Because of the consolidated and integrated description, with motivation and explanation, Dupin’s explanation of all of this, in sections 5, 6, 7, 8, may be additionally considered in our analysis. We have just noted the paragraphs that correspond to these sections. Section 5 begins with sentence 172 (in the set of 322 sentences). So the Dupin explanatory sub-narrative, in dialogue with the narrator, embraces sentences 172 to 322, that is, paragraphs 88 to 123. So the Dupin sub-narrative here comprises 151 sentences, that are in 36 paragraphs.

Our next step in data preprocessing is to select the word corpus that will be used. This starts with removal of all punctuation, numeric characters, and the setting of upper case to lower case.

It is reasonable here to remove tool words, also referred to as function words. In Murtagh [2005], chapter 5, and in Murtagh and Ganz [2015], the case is made for these function words in mapping emotional narrative or stylistics (e.g. to determine authorship), but these are not of direct and immediate relevance here. Instead, as outlined in section I, metaphor and metonymy are the forensic indicators, or perhaps even the forensic highlights, for us.

Sufficient usage of the word in the storyline is important. While very clearly the case that one-off (isolated, unique) use of a word can be very revealing, nonetheless we leave such an investigation to an alternative comparative study of storyline texts. Sufficiently frequent word usage both supports comparability between the text units we are studying, and also permits the focus of the analysis to be on inter-relationships, and not on uniqueness of word usage. Therefore we require the following for our word corpus: that a word be used at least 3 times in the overall storyline, and that this word be used in at least 3 of the text units (sentence, paragraph, section) that we are dealing with.

For the 322 sentences, we start with 1742 words. There are, in total, 7089 occurrences of these words. Then, having removed stop words, and requiring that a word appear in 3 sentences and be used at least 3 times, we find that our 322 sentences are characterized by 276 words. There are 1546 occurrences, in total, of the corpus of 276 words.

For the paragraphs, proceeding along the same lines, the 123-paragraph set is characterized by the 276 word set, and there are, as for the sentence set, 1546 occurrences, in total, of the corpus of 276 words.

For the sections, once again proceeding along the same lines, the 8-section set is characterized by the 276 word set, and, again clearly, there are 1546 occurrences, in total, of the corpus of 276 words.

This data preprocessing and selection is carried out for the following objectives.

Firstly, we will have one or more levels of text (hence, storyline) unit aggregation so that the principal factor space axes account for most of the information content. (Were it the case of having rare words in the analysis, then axes would be formed in the factor space to account for them.) We recall that for n text units, characterized by m terms, the factor space dimensionality will be $\min(n - 1, m - 1)$. This first point relates to the use of paragraphs and sections. (Let us note that in Bécue-Bertaut et al. [2014], where the flow and evolution of narrative is the aim, our aim is a little different here, because the text units that encompass the most basic text units, the sentences, can be themselves interpretable. Cf., e.g. vocal expression on a theme being all in one paragraph.)

Secondly, our selection of words directly impacts the interpretation of the data.

V A PRELIMINARY VISUALIZATION OF THE NARRATIVE STRUCTURE

We have here the successive sentences characterized by their constituent words, from the retained corpus. We firstly map the cloud of sentences, 322 sentence cloud in a 276-dimensional word set space, into a Correspondence Analysis factor space. Since the word set has been reduced from the original set of 1742 words, some sentences become empty. Non-empty sentences account for 310 of these 322 sentences. In Figure 5 we require words to be at least 5 characters long. This leads to a corpus of 205 words, with 293 sentences not becoming empty.

In Figure 5, sentences 11 and 12 are merged very early in the sequence of agglomerations, and these sentences are found to be quite exceptional. They are as follows: Narrator: “Nothing more in the assassination way I hope?”, Prefect: “Oh, no; nothing of that nature.” The two large clusters that are merged at the 3rd last agglomeration level have the last sentence of the first large cluster, and the first sentence of the second large cluster as follows. Sentence 182: “But he perpetually errs by being too deep or too shallow for the matter in hand; and many a school-boy is a better reasoner than he.” Sentence 193: “I knew one about eight years of age, whose success at guessing in the game of ‘even and odd’ attracted universal admiration.” This is early in what has been taken as the Dupin explanatory section of the narrative. The second very large cluster constitutes the major part of this Dupin explanatory part of the storyline.

In Bécue-Bertaut et al. [2014], it is described how the text units, taking account of the chronological order, can be statistically assessed (using a permutation-based statistical significance testing) at each agglomeration, for the agglomeration to be based on a pair of homogeneous clusters. This allows derivation of a partition of the set of text units. Since the chronological, hence contiguity, constraint applies, this partition is a statistically defined segmentation of the text units. In this particular work, we prefer to use paragraphs and sections, as described above, in view of their interpretability.

VI SEMANTIC ANALYSIS OF CHRONOLOGY USING STORYLINE SEGMENTS

In this section, we are most concerned with diachrony, or the evolution of the narrative. For this, we find a correspondence – what we may refer to as homology, in the sense of Bourdieu-related geometric data analysis – between a pattern that we uncover in the data, and Lacan’s registers, viz. the Real, the Imaginary and the Symbolic. In the storyline here, we find an evolution, or narrative trajectory, between these registers. Lacan’s registers are of value to us as an interpretive viewpoint.

It has been noted above, section I, how both synchrony and diachrony of the semantics of the storyline narrative are of importance here. As noted also, we can determine statistically a segmentation of the narrative. This is achieved through first mapping the narrative into the semantic factor space, taking account of all interrelationships of narrative text units and the words and terms that are associated with these text units. For interpretation, we prefer, see section IV, to use what we have selected as natural segments in the narrative text.

The cumulative percentages of inertia associated with factors 1 to 7 are as follows: 20.2, 38.1, 54.2, 68.8, 81.5, 92.9, 100. The principal factor plane is displayed in Figures 6, 7. The chronological trajectory is to be seen in the first of these figures. The second figure has a triangular pattern, that is a display of the narrative, with reference to the chronology of the narrative. Usually with such a triangular pattern, we look especially towards the apexes in order to understand it. Figure 7 shows the most important words.

From this display, taking the figures 6, 7 (not overlaid in the same figure, to make the displays clearer), we can conclude in this way: segments 1, 2, 3, 4 are gathering facts and impressions from the real; segment 5 advances into the imaginary; segment 6 expresses this in a symbolic way; and that allows a consolidated, integrated, “overall picture”, core of segments 7, 8.

In Figures 8 and 9, factors 3 and 4 are displayed. If this viewpoint expressed above is acceptable, namely that segments 5 and 6 comprise the move towards the Imaginary, then towards the Symbolic, then we can draw this perspective: that the effect of these two segments in the overall narrative is to take such segments as segments 3 and 4, operating in the Real, then work through the Imaginary and Symbolic discussion, and arrive then, as a consequence, at the final, terminal and more conclusive segments, segments 7 and 8. In a way, we are drawing the conclusion, from this particular storyline, as to how the Imaginary and the Symbolic serve to be taken into (and become part of) the Real, or how the Symbolic emerges from the Real and the Imaginary.

We will not pursue a very detailed study of further factors, mainly due to the following. Considering the important words that are in the first and second factor plane, Figure 7, it does problem solving related to school learning (cf. lower right quadrant), and problem solving related to mathematical reasoning (cf. upper right quadrant) are quite dominant themes.

Moving on now the third and fourth factor plane, Figure 9, a more interesting perspective, given our interests in this work, is revealed. We propose the following perspective on this figure, Figure 9. Take the words on the left, negative half axis of factor 3, as pertaining to a real register. Therefore, mostly, they betoken the unknown or the unknowable. Next, take many of the words displayed in the upper right quadrant as associated with the imaginary. This includes “furniture”, “houses”, “microscope” and so on. This is how we can imagine problem-solving. Thirdly, and finally, take many of the words in the lower right quadrant as betokening the symbolic. What we have here is money, payment. In other words, in a practical setting here, the problem solving is associated with the symbolic value of money.

We conclude that Lacan’s registers have been of major benefit in providing semantic-related understanding of the essential pattern that we determined in the narrative chronology. Such homology of semantic structure, i.e. morphology of narrative, is, in our view, to be sought in any domain that can be modelled through Lacan’s registers.

VII SEARCHING FOR SYNONYMS

Using our active core set of 8 narrative segments, crossed by the corpus of 276 words, with 1546 occurrences of these words, we look at a partition of this word set. Using factor projection coordinates, endowed with the Euclidean distance, and in the context of this work, the full dimensionality of 7 dimensions, allows us to construct a hierarchical clustering of the word set. The minimum variance or Ward agglomerative criterion is used, in harmony with the inertia of the cloud of segments, and the cloud of words (with identical inertia for these clouds of profiles of narrative segments, and of word profiles). The most significant cut of the dendrogram resulted in 4 clusters in the resulting partition. For each of these clusters, we can read off the closest words to the cluster centre, and the members of the cluster that are most distant from all other clusters. In order to have some insightful clustered collections of words, we used a lower level cut of the dendrogram to produce 20 clusters in the following, and here we report on the following clusters. Provided are both the words that are closest to the cluster centre, and the cluster member words that are furthest from other clusters.

1. Cluster 3, concerning the symbolism of mathematics terms. We note the links with failure to find, and concealment.
 “concealment”, “given”, “hidden”, “identification”, “ingenuity”
 “action”, “algebra”, “depends”, “error”, “fail”
2. Cluster 1, relating to the school and marbles, we would claim – imaginary – part of the narrative. It expresses how bad behaviour can win out.
 “school”, “second”, “even”, “boy”, “cunning”
 “guess”, “simpleton”, “wins”, “second”, “kind”
3. Cluster 4, concerning how being Parisian is played out here.
 “therefore”, “simple”, “odd”, “parisian”
 “simple”, “parisian”, “therefore”, “odd”
4. Cluster 16: concerning certain aspects of the role of power.
 “fool”, “possession”, “personage”, “power”, “although”
 “ascendancy”, “still”, “robber”, “months”, “power”
5. Cluster 19: concerning specific domains of search.
 “microscope”, “every”, “article”, “examined”, “removed”
 “furniture”, “houses”, “probed”, “accurate”, “entire”

From looking at the close semantic (i.e., based on the semantic factor space embedding) association of cluster members, we have pointers to what could play the role of metaphor, being locally and temporally, contextualized synonyms.

VIII SYNCHRONOUS, SEMANTICALLY-BASED CHAIN OF WORDS

Let us consider a chain that spans the entire word set. While this may be locally defined from sub-narratives, here we will consider the Hamiltonian path, that will be based on a single link hierarchy. Hence the minimal distance between clusters determines the links in this comprehensive, all encompassing, chain.

In the ordering of words that is associated with the single link hierarchy, the immediate words before and after “letter” are as follows:

“admeasurement”, “observation”, “form”, “surface”, “**letter**”, “search”, “asked”, “course”, “premises”.

We suggest that the earlier words here are somewhat imaginary, that the word “letter” is symbolic in a major sense in the Poe story, and that the words that follow are indicative of the real.

IX STATISTICALLY SIGNIFICANT WORD ASSOCIATIONS

For close associations of words leading to either metaphor or metonymy, we adopt the following principles. Firstly, we seek such associations from the data, and not to impose an a priori statistically-based probabilistic model or other prespecified criterion. Secondly we want to have such associations contextualized. The latter is for the seeking of associations to be in semantically-defined clusters.

In this sense, we are looking for pointers towards the triad that defines a metaphor. Consider how Ricoeur (Ricoeur [1977], p. 276) conceptualized this: “We arrive at metaphor in the midst of examples where it is said, for instance, that a certain picture that *possesses* the colour grey expresses sadness. In other words, metaphor concerns an inverted operation of reference plus an operation of transference. Close attention must be paid, therefore, to this series – reversed

reference, exemplification, (literal) possession of a predicate, expression as metaphorical possession of non-verbal predicates (e.g. a sad colour).” Thus in brief, we may consider here that $x = \text{picture}$, $y = \text{grey}$, $z = \text{sadness}$, and we have the proximity of x and y that we may view as comprising the apexes of the base of an isosceles triangle.

Let us again describe our data: 322 sentences; structured as 123 paragraphs. All paragraphs are associated with the speaker, or discussion of an individual, so that name is, or can be, associated with each paragraph.

We extract words that are of minimum length 1, assigning all upper case to lower case, removing all numbers and punctuation. For the 322 sentences, this provided a corpus of 1742 words. The total number of words in the Poe novel, at this level of the processing, was 7089. A Correspondence Analysis was carried out on this data, 322 sentences crossed by the corpus of 1742 words. A limitation to have at least three presences of use of a word, and that the word be used in at least three sentences, was placed on occurrences of words for two reasons: comparability of sentences is based on at least some shared words; and very rare word occurrences could over-imbalance the outcomes. Also, R package *tm* stopwords were removed, as well as a set of our user defined keywords (14 of these including e.g. “s”, “for”, “also”, and others).

The resultant sentence set was crossed with a corpus of 276 words (reduced from 1742), with, in total, now 1546 word occurrences. Due to the removal of words, stopwords, and rarely used words, some sentences became empty. Removing these resulted in 310 sentences being retained. Percentages of inertia for the first few axes were as follows: 1.25, 1.23, 1.22, 1.19, 1.17. Although small in value, nonetheless we are most interested in their prioritisation of information.

We also investigated the chronology based on the following: the sequence of sentences; the sequence of paragraphs, i.e. text segments, that were mostly either a continuous speech segment, or relating to an individual; a set of eight sections covering the entire story that was manually segmented, approximately in line with the timeline; and four statistical segmentations of the storyline based on combinatorial probabilistic significance levels. Successive segmentation of the storyline was, respectively, with the following numbers of segments: 322, 123, 8, 46, 26, 13, 11. It was found that these sequences were weakly correlated with the factors. As supplementary elements on the factor space planar projection, they were very close to the origin. We conclude that there is not much that carries chronological meaning in this story. That is on the global or overall level. Word associations or sequences (that could play a role in metaphor or in metonymy formation) are a different issue.

Based on the Correspondence Analysis factor space mapping, endowed with the Euclidean distance, the clustering of sentences and also of words was investigated. Although distinct in regard to the basis for the clustering, while of course using the minimum variance – hence inertia in the Euclidean-endowed factor space – agglomerative criterion, the outcomes implicitly share the 5-dimensional (used just by default as a small set of factors) input.

It has already been noted how factor 1 counterposes the specifics of investigation to the ancillary small sub-narratives, relating to mathematical thinking analogies (upper right quadrant) and to the schoolchild motivation and decision-making analogies (lower right quadrant).

The 5-class partition obtained allows us to look closely at some of the clusters. These clusters are of cardinalities, for the words: 10, 218, 5, 19, 24, and for the sentences: 8, 258, 6, 21, 17. They are in sequence of their mean value projections, from left to right on the first axis.

Let us look at low level partitions in the dendrogram in order to select small cardinality, very compact clusters. Following Husson et al. [2011] (p. 151) we can use the v-test of association of the category presence values relative to the mean value of that variable. This allows for a null hypothesis test of “the average ... for [the] category ... is equal to the general average”, “in other words, [the] variable does not characterise [the] category ... and can therefore calculate a p-value. A p-value not far from zero indicated rejection of that null hypothesis. That is to say, a p-value near zero indicates that the variable emphatically does characterise the category.

When we look at an 11-class partition we find class 1 consisting of:

```
Class 1:
          p.value of H0 using v.test
puff      1.188185e-13
abernethy 1.097031e-03
```

```
Class 2:
          p.value for v.test
probed    1.693836e-06
looked    1.758749e-03
```

Class 5 with the following words, with p-values of the v-test less than 0.05 (therefore rejecting the null hypothesis here at the 95% significance level): letter, man, ordinary, gname, reward, asked (Here gname is the Prefect. There is for example the following in the Poe text: “Monsieur G -- , the Prefect of the Parisian police.”). In this 11-cluster partition, class 10 is mainly about the mathematical analogies, and class 11 is about the schoolboy analogies.

In order to find some small clusters, leading to usefully relationships that are semantically very close due to cluster compactness, we looked at various sized partitions derived from the hierarchical clustering dendrogram. From a 50-cluster partition, we find the following.

Cluster 40 consisted of the words “mathematician”, “poet”.

Cluster 43 consisted of the words “example”, “analysis”, “algebra”.

Cluster 47 consisted of the words “reason”, “mathematical”.

Cluster 48 consisted of the words “truths”, “general”.

Cluster 50 consisted of the words “truths”, “mathematical”.

Cluster 15, including “letter” had these words: “possession”, “letter”, “premises”, “still”, “since”, “observed”, “said”, “main”, “far”, “power”.

Cluster 1 consisted of “puff”, “abernathy”.

Cluster 20 consisted of the words “document”, “especially”, “things”, “point”, “importance”.

Cluster 27 consisted of the words “personage”, “document”, “royal”, “thorough”, “necessity”, “question”, “make”.

Our overall objectives here are to determine potentially interesting word associations, that could then be taken as, or found to be, some triadic metaphor (synchronic) relationship, or metonymy, a diachronic relationship. In very general analogy to the observational science of astronomy, we do not seek to statistically test the properties of what is found, but rather to obtain relevant, candidate relationships, that, as candidate relationships, will then be assessed further in other

contexts. Such, we may wish to state, could be considered for the words “poet” and “mathematician” in this case.

X CONCLUSIONS

Through the semantic mapping of the storyline, it has been described how patterns found can be related to Lacanian registers. This has been covered in sections leading to section VI.

Then, as the basis for detecting and observing synchronicity-based metaphor, and diachrony-based metonymy, some related approaches have been described and exemplified in sections VII to IX.

In the sense of unsupervised classification and exploratory data analysis, our approach is both “The model should follow the data, and not the reverse!” (Benzécri quotation) and “Let the data speak for themselves” (Tukey quotation).

Our text analysis has pointed out the intertwining among three Lacanian registers. Indeed, in Figure 8, the storyline segments of semantic analysis, identify a quasi-cyclic circuit starting from the Real and the Imaginary registers to the Symbolic one. For instance, considering Figure 9, the semantic storyline starts from a cluster in which there are words referring to Real, like “dark”, “affair”, to go toward a second cluster in which there are terms like “probed”, “removed”, “escape”, still belonging to the Real, all referring to the dramatic dimension of the Real as inherent in the potentially compromising content of the purloined letter. Hence, this storyline goes on toward Imaginary, crossing a cluster whose terms – like Dupin – belong to the Imaginary, to ending in a cluster containing a set of words belonging to Symbolic, like “paper”, “document”, “possession”, “personae”, “card”.

Likewise, in the case of factors 5 and 6, we find that the semantic storyline starts from a cluster containing terms belonging to the Real, like “affair”, “dark”, “mystery”, to go toward a second cluster containing words still belonging to the Real, like “really”, “check”, “puff”. Then, this storyline goes to a cluster whose terms belong to the Imaginary, like “Dupin”, “said”, “known”, to finish in the Symbolic with terms of the type “card”, “seal”, “rack”, “cipher”.

In any case, in all the planar projection plots related to this semantic analysis of chronology by means of storyline segments, we note that Imaginary clusters are almost always placed in the centre of each diagram (clearly in Figure 8), besides to be the intermediate, hinge step between the Real (the realm of angst and fear according to Schmitz) and the Symbolic (socio-symbolic domination of Schmitz). So the Symbolic roughly corresponds to Schmitz’s Habitus-field intermezzo, coherently with the fact that Lacan Symbolic corresponds to Freud’s Super-Ego agency, the place in which there takes place the crucial passage from thing representation to word representation. Furthermore, we also note the prevalence of Real register in the first steps of semantic storyline, moving to Imaginary toward Symbolic, the prevalence of unconscious realm underlying conscious meaning of language.

In conclusion, our Correspondence Analysis of Poe’s story has been useful in identifying certain formal structures resembling Lacan registers action in giving rise to language.

References

- M. Bécue-Bertaut, B. Kostov, A. Morin, and G. Naro. Rhetorical strategy in forensic speeches: Multidimensional statistics-based methodology. *Journal of Classification*, 31:85–106, 2014.
- J.P. Benzécri. *L'Analyse des Données, Tome II Correspondances*. Dunod, Paris, 1973.
- J.P. Benzécri. *Histoire et Préhistoire de l'Analyse des Données*. Dunod, Paris, 1982.
- J.P. Benzécri. L'avenir de l'analyse des données. *Behaviormetrika*, 14:1–11, 1983.
- M. Grenfell and F. Lebaron. *Bourdieu and Data Analysis: Methodological Principles and Practice*. Peter Lang, Bern, 2014.
- F. Husson, S. Lè, and J. Pagès. *Exploratory Multivariate Analysis by Example Using R*. Chapman and Hall/CRC, 2011.
- B. Le Roux and H. Rouanet. *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*. Kluwer (Springer), Dordrecht, 2004.
- F. Lebaron and B. Le Roux. *La Méthodologie de Pierre Bourdieu en Action: Espace Culturel, Espace Social et Analyse des Données*. Dunod, Paris, 2015.
- P. Legendre and L. Legendre. *Numerical Ecology*. Elsevier, Amsterdam, 3rd edition, 2012.
- F. Murtagh. *Correspondence Analysis and Data Coding with Java and R*. Chapman & Hall, Boca Raton FL, 2005.
- F. Murtagh and A. Ganz. Pattern recognition in narrative: Tracking emotional expression in context. *Journal of Data Mining and Digital Humanities*, 2015, 2015.
- F. Murtagh, A. Ganz, and S. McKie. The structure of narrative: The case of film scripts. *Pattern Recognition*, 42: 302–312, 2009.
- E.A. Poe. The purloined letter. URL <http://americanliterature.com/author/edgar-allan-poe/short-story/the-purloined-letter>.
- E.A. Poe. *Tales by Edgar A. Poe*. Wiley and Putnam, New York, 1845. URL <https://ia600408.us.archive.org/0/items/tales00poe/tales00poe.pdf>. pages 200–218.
- E.A. Poe. The gift, christmas, new year and birth present. 1845 (MDCCCXLV). URL <https://ia802706.us.archive.org/2/items/giftchristmasnew00carerich/giftchristmasnew00carerich.pdf>. Pp. 41–61.
- E.A. Poe. Edgar allan poe, the purloined letter, 2016. URL http://www.eng.fju.edu.tw/Literary_Criticism/structuralism/purloined.html. English Language Literary Criticism, Fu Jen Catholic University, Taiwan.
- P. Ricoeur. *The Rule of Metaphor*. Routledge, London GB, 1977.
- A. Schmitz. The space of Angst, 2015. presentation (40 slides), Empirical Investigation of Social Space II Conference, Bonn, Germany, 12–14 October.
- A. Schmitz and M. Bayer. Strukturele Psychologie: Konzeptionelle Überlegungen und empirische Analysen zum Verhältnis von Habitus und Psyche (Structural psychology: Conceptual considerations and empirical analyses on the relationship of habitus and psyche), 2014. preprint, 17 pp.

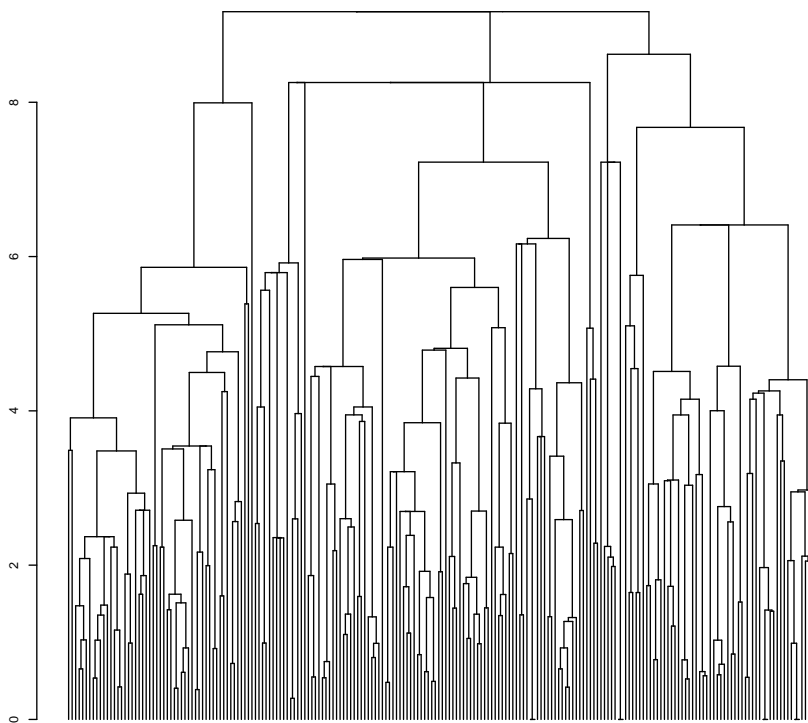


Figure 1: Hierarchical clustering of sentence by sentence, based on the story's sequential structure. There are 213 sentences here, being the terminal (or leaf) nodes, ordered from left to right. Each sentence contains some occurrences from the corpus of 48 nouns that are used.

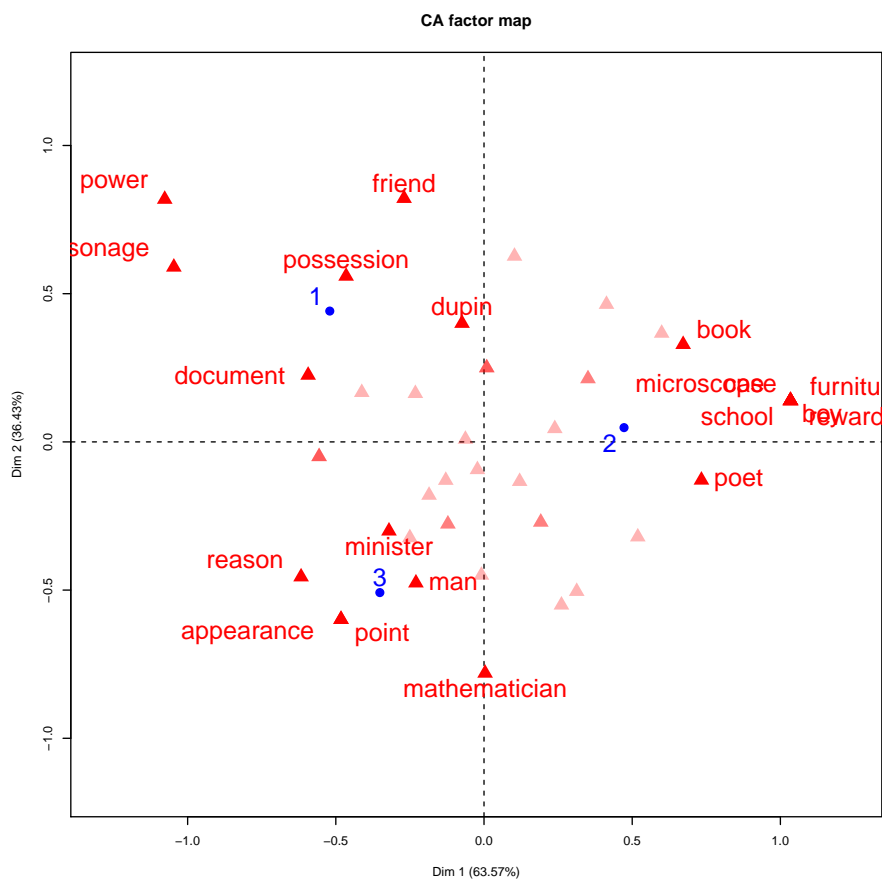


Figure 2: Correspondence Analysis, top contributing 20 words. Words with high contribution, somewhat overlapping in this display, with projections on the positive factor 1 are: case, microscope, school, reward, boy, furniture; and book, poet. Also displayed are the three acts, 1, 2, 3.

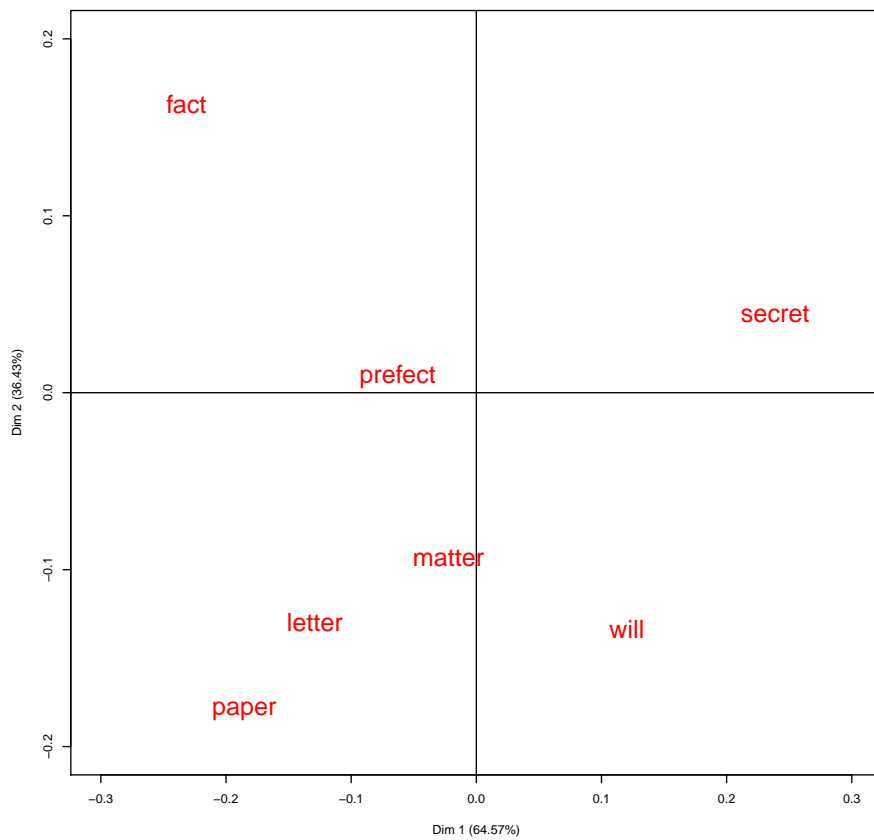


Figure 3: From Figure 2, here are shown the words that are near the origin.

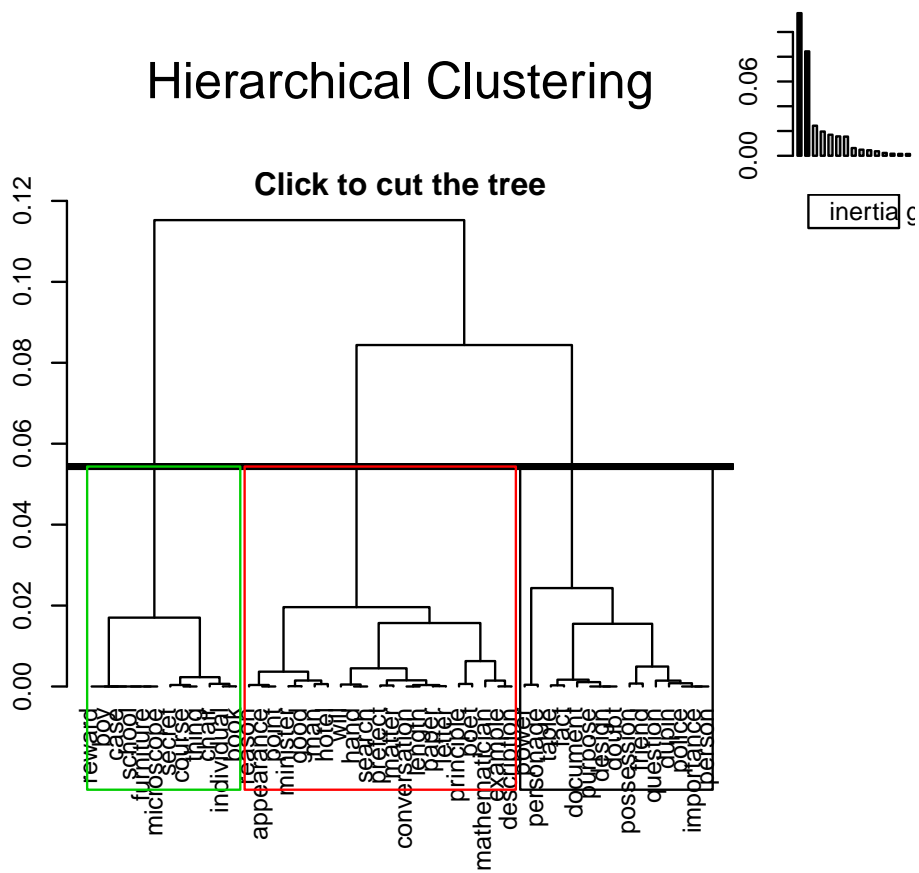


Figure 4: Hierarchical clustering of the word set, from the same 5-axes, factor space, semantic mapping of this data. CA, principal plane of acts crossed by retained nouns table.

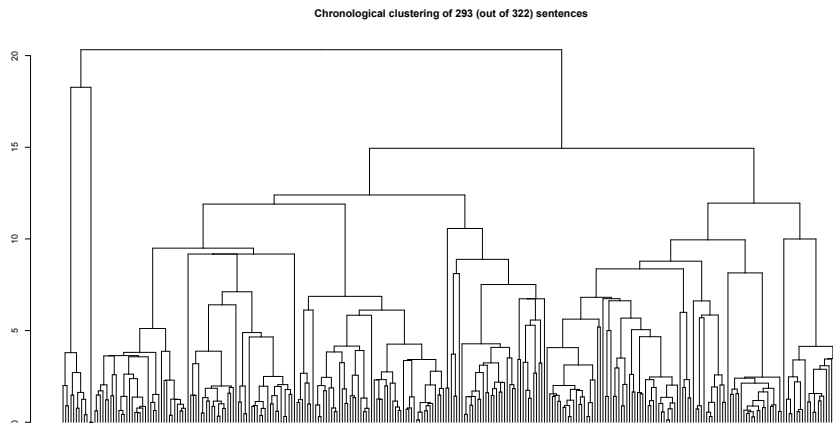


Figure 5: Contiguity-constrained, where contiguity is the chronology or timeline, hierarchical clustering of the 322 sentences. These sentences are characterized by their word set (1087 occurrences of 205 words). This hierarchy is constructed in the factor space, of dimension 5, that is endowed with the Euclidean metric. Due to the reduced word entailing that some sentences become empty, the number of sentences in the correspondence factor analysis was 293 (from the 322).

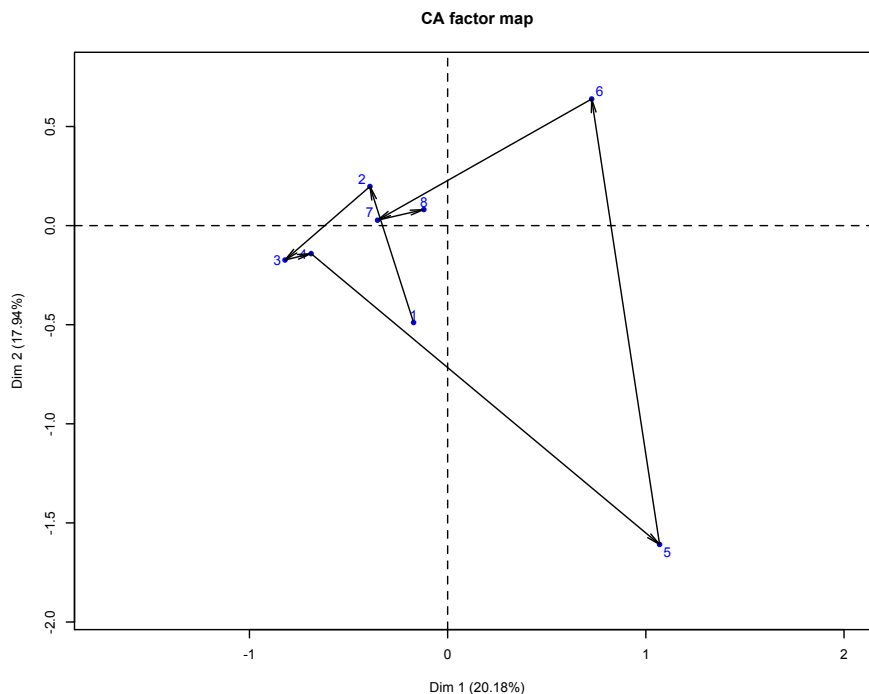


Figure 6: Principal factor plane of the 8 story segments crossed by the 1546 occurrences from the selected 276-word corpus. Arrows link the successive segments, numbered 1 to 8.

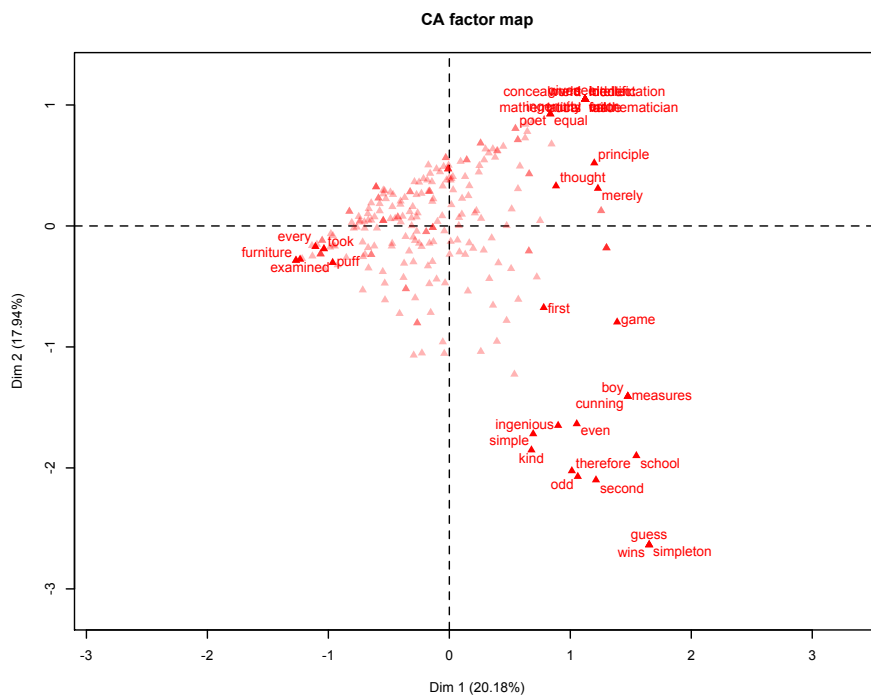


Figure 7: Displayed here are the 40 words that most contribute to the inertia of these factors, factors 1 and 2. In the upper right (beyond equal, poet), terms are: mathematician, world, value, truths, see, mathematical, intellect, fail, error, ingenuity, identification, hidden, given, concealment, reason, hand.

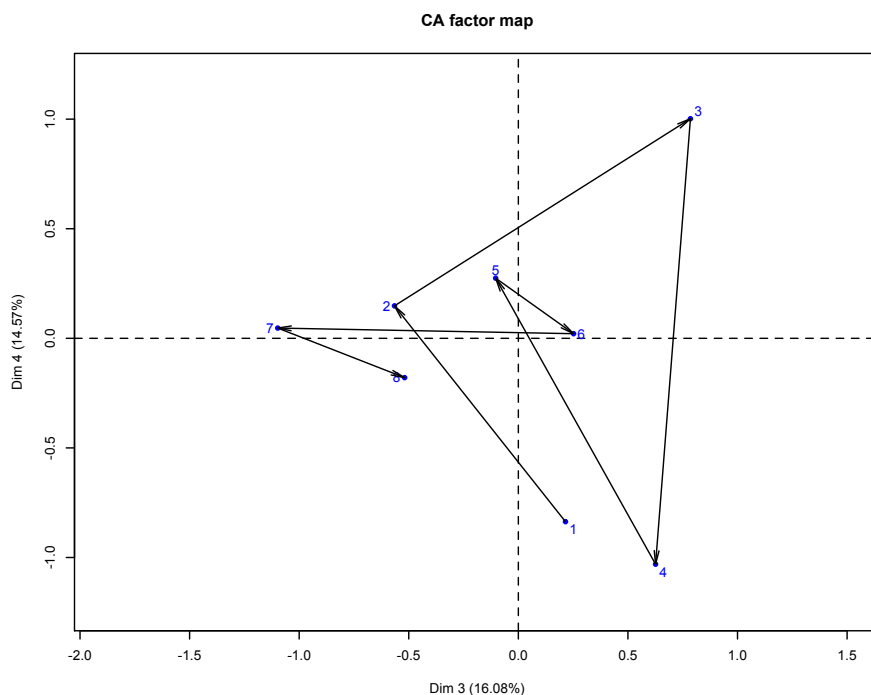


Figure 8: Plane of factors 3, 4, displaying the 8 story segments, with arrows linking the successive segments, numbered 1 to 8.

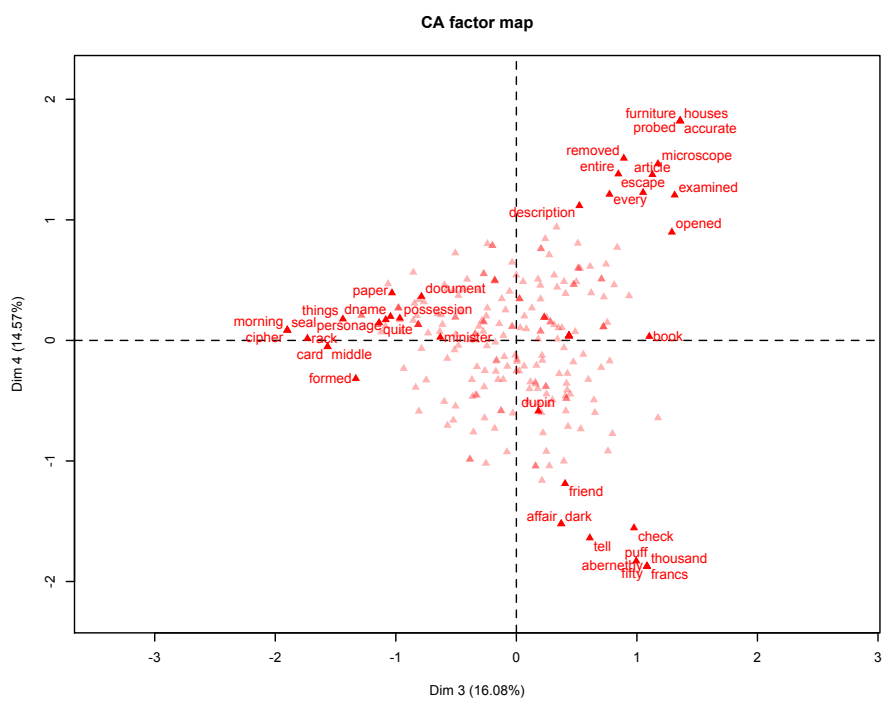


Figure 9: The 40 words that most contribute to the inertia of these factors, factors 3 and 4. The word dname is a rewritten form of “D–”, i.e. Minister D.