Text analysis: an introductory manifesto

Book section

SAGE Benchmarks in Social Research (4 volumes)

TEXT ANALYSIS

Text Analysis – An Introductory Manifesto

Martin W Bauer, Ahmet Süerdem, Aude Bicquelet

1 A working definition of 'text' for social science analysis

> *...the discourse on the Text should itself be nothing other than text, research, textual activity, since the Text is that social space which leaves no language safe, outside, nor any subject of the enunciation in position as judge, master, analyst, confessor, decoder. The theory of the Text can coincide only with a practice of writing.' (Barthes. 1971)*

Selecting the articles for these volumes of SAGE benchmarks on 'text analysis' was no easy task. How to determine the scope of the selection? One could go with a very limited definition of text, such as a canon of official documents or a very broad notion, like 'cultural artefacts', representing any meaningful symbol system. These definitions of text resonate with different approaches to text: decoding and deconstruction. The canon selection suggests that the meaning of a text is closed, contained in the work with the sole purpose to transmit a message from author to reader. Within a 'transfer-conduit' perspective (see Reddy, 1993), the aim of text analysis is to provide expert tools such as literary criticism, philology, or content analysis to decode the texts which would otherwise be inaccessible for a simple reader; text analysis aims to observe and discover the attitudes, behaviours, concerns, motivations and culture of the text producer from an expert point of view. According to the open definition on the other hand, the meaning of any artefact, including text, is wide open, the message is not there to discover and to deconstruct during the reading process. Recovering the meaning is

not an exoteric activity (for experts and the educated), but an esoteric performance (immersive and emergent). But, reading is an interpretive activity that can only be performed by those who are embedded into the symbolic world of the text. All action, if we push the notion, even nature, is a "text" to be read, where signs are intelligently designed to reveal knowledge and guide the way to truth. The purpose of text analysis is thus not the passive reading of the author's world but the entry into a reflexive dialogue between the reader-analyst and the text.

Our definition of text analysis straddles the space between the two extremes of 'decoding' and 'deconstruction'. A social scientific text analysis aims *to explain the life-world within which the text is embedded; to open up the perspective of the author that is delineated by his/her social and cultural context and to draw attention to the structural aspect of everyday practices and meaning patterns.* Yet, the position of text analyst as a reader should avoid a "judge, teacher, analyst, confessor, or decoder" role. To analyse a particular text is also to produce it, a self-reflexive activity providing readers with insight about the life-worlds of others, a phenomenological exercise for comparing one's lived experiences with those of others, modifying one's perception of the world and coming to a common, inter-subjective construction of social reality by fusing horizons that were hitherto separate.

In this sense, text analysis stands on the principles as qualitative research as defined by Flick et al. (2004, p7):

- *Social reality is understood as a shared product which makes sense to the members of a community.*
- *This sense is not a fact to be discovered, but an unfolding reflexive process.*
- *'Objective' circumstances are relevant to a life-world through subjective meanings.*
- *The communicative nature of social reality permits the reconstruction of constructions of social reality to become the starting point for research."*

This definition approximates text analysis with qualitative research. According to Geertz (1973) and Riceour (1973), social action can and should be read as text; text is the model of social life. Studying social life does not discover universal laws of human behaviour, a ambition often characterised as 'physics envy', but involves interpreting social life within the

variable framework of symbol systems. The social sciences' primary purpose is not prediction of human behaviour as the physical sciences do for the movement of objects. Social research is first of all the reading of social actions, i.e. understanding, explaining and interpreting actions to render them intelligible through inter-subjective meaning. This might be the difference between human movement considered as 'behaviour' or as 'action'. And reading an action is a discursive activity, not simply describing but also making a statement about the desired state of the world. To call an act of violence 'terrorism' is more than just a neutral word; it is a call for action against those who are called 'terrorists'. Social analysis itself is discursive and involves more than presenting a body of facts. Reading social actions and writing up the research is a discursive act. Accounts of data analysis are narrative constructions and they must be treated as combinations of fact and fictions. They are valid and significant if their rhetoric is persuasive and makes sense. There is no *p*-value or fit–statistic to benchmark 'making sense'; fit-statistics are part of the rhetoric of credibility.

However, despite convergence we can distinguish textual analysis and qualitative research in terms of their sources. Qualitative research traditionally recognises three sources of empirical data: **interviewing, observing and documents**. Interviewing involves listening skills and the conversation may be voice recorded, and later transcribed into a text stream. Qualitative researchers also observe and personally witness what people are doing, how they deal with themselves, things and other people. These observations are often transformed into text formats. Interviewing and observation can be distinguished from documents because they are face-to-face and thus obtrusive; they are produced for the purposes of the research and interviewer and observer effects need to be considered. On the other hand, documents are usually produced independently of the present researcher in a naturalistic environment (see Webb et al. 1966).

Documents are diverse, but their common feature is that they are left-overs of some kind of activity; they are produced in one context and used by the researcher in a different one. For example, while press news informs the readers on current affairs, they also offer the remote social scientist insights into social practices and narratives (who, where, when, what, and with whom) of a society and an epoch. Documents open up sources of information where data would otherwise be hard to come by because of spatial or temporal distances. Introduced by historians as witness evidence onto a distant past, the use of documents is now widespread across many domains of social sciences. We limit the scope of text analysis to the analysis of documents, although all social data are textual on one or other form.

However, the premise that documents are produced in naturalistic environments should not suggest that they can be treated as ‘more objective’ sources of data than other formats. Although documents are produced outside the specific research purposes of later years, their production, selection and analysis are not independent from thoughts, feelings, ideas, beliefs and intentions of social actors. First of all, documents are produced by individuals who communicate a mode of thinking. Second, they are often produced to give a justificatory account; thus the mind-set of an audience is implicit and rhetorically anticipated. Third, text analysis itself starts with preconceptions that are bounded by the socio-historical context in which it is performed; the mind-set of the analyst frames the data. We must not reify documents in and of themselves as ‘more objective’ data; they are facts constructed by the intervention of the researcher who selects them into a corpus and interprets them.

However, the interpretative nature of text analysis does not necessarily suggest that the analytic process should always be entirely subjective. Texts are produced within an institutionalised context of writing and action. Authoring is not an individual act but claims ‘authority’ to speak on someone’s behalf; the ‘Zeitgeist’ (the mentality of time and place) speaks through the author. Texts represent values, beliefs, rituals and practices of a community. And this repertoire of coded signs maps out the life-world of members of that community (see Bauer & Gaskell, 2008). Meaning does not reside in static and self-contained units but is constructed as a distributive, dynamic and inter-subjective performance involving contesting, negotiation and different understandings. This performance occurs within a semiosphere where sign repertoires are interwoven with layers of life-worlds (Lotman, 2005). In that respect, systematic analysis of texts gives us important clues about the historical and social conditions of the context within which they are produced.

In a nutshell, our understanding of analysing text involves reading any artefacts ‘showing designed texture’ of a symbol system and reflecting regularities in social practices. However, we hesitate to extend this by way of metaphor to understanding the ‘world as text’, the cosmos as a message, or the book of Nature. We stick to the restricted definition of ‘text’ as composed written material for operational purposes (Segre & Kemeny, 1988, 300ff). Our working definition marks some immediate exclusion: we will be dealing neither with sound nor with image materials as ‘text’. Although these modalities produce equally useful data streams for the social sciences (Bauer & Gaskell, 2008); they are better examined as a separate domain of inquiry. These volumes limit the scope of analysis to written documents.

This scope of the exercise highlights that papers considered in this collection accept texts as artefacts a) designed with a purpose, b) written in a natural language, c) produced in a genre with basic rules of production and d) which may help us to inter-subjectively reconstruct the life-worlds of producers and audiences of texts within a context. We are not limited to formal contexts although this might designate texts with more authority. Our definition of text involves the authority of all voices: it treats everyday texts such as personal diaries or newsprint in the same manner as literary works, legal statutes or Holy Scripture.

2. Complementary and overlapping SAGE collections

In collecting key papers for these SAGE volumes on text analysis we inevitably faced the issue of demarcation from and overlap with other projects in this series. We sought to achieve a complementary perspective without reproducing or replacing any existing collections. We identified several volumes in the SAGE Benchmarks Series where we could have found overlapping concerns, concept and citations.

- Atkinson P & S Delamont (2010) Qualitative Research Methods, London, SAGE
- Drew P & J Heritage (2006) Conversation Analysis, London SAGE
- Hansen A (2009) Mass Communication Research Methods, London, SAGE
- Hutchby I (2008) Methods in language and social interaction, London, SAGE
- Franzosi R (2008) Content analysis, London, SAGE
- Prior L (2011) Using documents and records in social research, London, SAGE

The overlap is least with the volumes by Drew & Heritage (2006) and Hutchby (2008). Both series deal with the analysis of verbal interaction, and in the very specialist manner of the pragmatics and socio-linguistics of conversations. We expect some overlap with the volumes of Atkinson & Delamont (2010), not least as their concern spans the entire field of qualitative research, in which textual data figures large. The overlap is probably larger with the volumes edited by Hansen (2009), by Prior (2011) and by Franzosi (2008). With Hansen we share an interest in mass media contents. For social science text analysis, the mass media are indeed a major data source, both for method development and as a field of substantive research. Equally, we share common ground with Prior (2011) on text documents; however, for Prior's edition the critique and analysis of the strategic contexts of text production is the key concern. Our present collection will have most overlap, conceptually and in selected papers,

with Franzosi's (2008) volumes on Content Analysis. It is therefore necessary to say a few words on how we see the difference between content analysis (CA) and text analysis (TA). We will return to this issue below.

3 Language confusions in the text analysis community

One of the difficulties of text analysis in the social sciences is the Babylonian confusion over terminology for text elements and analytic operations. Text analysis has been developed by different, sometimes distant, disciplines each having their own language game.

For example, consider the **philological studies** of canonical documents. Here a sophisticated methodology has developed to secure the 'true' version of a text underneath a myriad of versions and translations, and to validate interpretations with historical, dogmatic or literal methods. **Literary criticism** has developed analytic categories arising from different traditions such as hermeneutics, semiotics, de-construction and reception studies for the purposes of interpreting the meaning of a literary work. **Linguistics**, with much concern for syntax and style brings structuralist language analysis to the game. **Artificial intelligence**, focussed on simulating natural language processing, is creating text mining routines and automatic pattern detection for extracting and analysing text corpora from text streams such as social media. **Historical studies** have elaborated the critical approach to examine documents and to distinguish the fake from the genuine article in order to reconstruct credible historical testimony. **The social sciences** have developed their own terminology around sampling, coding, framing and thematic organisation and statistical analysis. Each of these fields of enquiry is highly specialised and pays little attention to the neighbouring pursuits, thus language spills over into this grand confusion.

Take just the simple example of using words like tagging, coding, indexing or mark-up. Do we use these words of different origins ('tagging' = linguistics; 'coding' = social science, 'indexing' = philology and library sciences, 'mark-up' = computer science) interchangeably or do they serve to identity different things? The confusion arising when text researchers talk of coding and mean indexing or tagging seems small, but is a pressing issue in the teaching of text analysis. Other confusions involve epistemological posturing over issues of induction, deduction, abduction, positivism, phenomenology and constructivism.

One might take a pragmatic position and agree that these are only matters of words, little to worry about, as long as the researcher is served. If I mark-up a text, and call it meta-data or

indexing, who cares? If another calls this tagging or coding, so be it. However, language matters as it determines the way we carve up the world. A clearer convention of text analytic concepts and operations is desirable to sort out the key terms from different traditions of dealing with texts. However, this effort is as much about raising awareness as it is about offering final definitions, as policing the text analysis language is not our intention.

4 Key dimensions

In collating key readings on text analysis, we felt that three criss-crossing tensions beset many of the discussions, either implicit or explicit. These tensions throw a light on some of the debates and polemical positioning arising in TA.

- Reading versus using a text
- Structural analysis versus interpretation
- Qualitative and quantitative approaches

Reading versus using a text

Reading a text refers to activities that and focus on empathy and understanding the life-worlds of others, be that the author, the text structure or the audience of reception, and the wider context of writing and reading. This is non-instrumental reading for reading's sake. Reading celebrates the possibility of transformative experiences: the reader is changing themselves through an 'aesthetic' encounter with the other. Reading Dostoyevsky's 'The Brothers Karamazov' can turn you into a different person touched by the events and characters. Reading opens the possibility that something unexpected is happening, frustrates a prejudice, brings a new understanding through the 'fusion of horizons', your own and that presented by the book. Reading means entering into a dialogue with another person. The text indeed re-presents and thus gives voice to the one(s) that authored the text. But, reading culminates in a reflexive, reconstructive act where the text and reader jointly reach out to something new.

However, the deconstructive idea of *différance* (Derrida, 1967*)* or infinite regress to a referent may reduce the interpretive process to a vicious circle of speculative language games. When the analyst is free to pursue their own rhetoric of interpreting the text without

any constraints, then the interpretive process can easily morph into demagogy where creating "over-interpretations" is the game of the day (Eco, 1992). Hermeneutic processes must take preconceptions as a starting point. Interpretation is not free from historically effected consciousness (Gadamer, 2004). To escape eisegesis, i.e. imposing one's own agenda on the text, one must not deny prejudice and then be caught up in it, but be critically reflexive of prejudice and acknowledge it in order to gain novel insights. The hermeneutic circle becomes a productive exegesis through an iterative process of critically examining the cultural prejudices of the author, the text, the original audience, and the analyst themselves.

In contrast to all this, **Using a text**[1] refers to activities that make TA an instrumental activity for purposes other than understanding the text. For example, we might use texts as convenient indicators of something outside the text such as cultural or social structure. This is also called 'symptomatological reading'. Text elements are treated as if they were symptoms of hidden processes, like a fever is the symptom of the body fighting an infection. We might compare the vocabulary of different texts as indicators of social class positions, or grammatical feature changes as indicators of social change. An extreme example is the recent launch of Google Trends, where we are invited to sift through millions of online documents to get an instant indicator of the changing prevalence of keywords, while access to the original text is not possible. Reading is no longer part of this operation of machine search-and-retrieval. Probably most of classical content analysis falls into this category; with the coding we cut the link to the original document; the code represents the document for all future purposes. Although the technological trend seems to suggest that reading is less important, an interpretive turn in the social sciences might however strengthen the awareness of this contrast between reading and using a text. The interpretive activity might reassert itself, and we consider our present collection of key papers on text analysis as a balance of both trends.

Structural analysis versus interpretation

The second dimension we want to consider is between focus on structural features versus focus on interpretation.

[1] Note that this definition of 'using a text' must not be confused with pragmatics, which is concerned with the practical use of signs in everyday life and which is indeed an interpretive activity.

The **analysis of structural features** considers the text merely as a design. The signs making up a text are organised sequentially (syntagma) and selected from a system of replacements (paradigm) to signify. The words form a vocabulary that can be assigned into grammatical-functional categories such as subjects, objects, verbs. These categories form sentences according to syntactic rule, and words form semantic relations by appearing in the same context. Finally, text displays sequential order, style and discourse, an order that is recognisable above the level of the sentence. All in all, the structure of a text is an organised taxonomy of linguistic resources that may be arranged into meaningful configurations. The syntax constrains these permutating configurations. Signs do not signify except in their reference to other signs.

However, generating meaning in terms of signs referring to other signs is problematic as Eco (1990b) underlines: "The meaning of a representation can be nothing but a representation ... the interpretant is nothing but another representation ... and as representation, it has its interpretant again. Lo, another infinite series" (pp 28). The interpretant of a sign "becom(es) in turn a sign, and so on *ad infinitum*..." (pp 35-6). Now you can spend all your time analysing words and sentences or paragraphs, counting and comparing with others, without worrying much about what is being said, with no need to understand anything. You might come across a linguist who declares with pride that they study the syntactical structures of Dutch in comparison to Nepalese, without being able to understand a conversation in any of these languages.

While such a structural analysis might translate a text from one language to the other like ‘Google translate’ does; or pass the Turing test and simulate a human chat, predict the next sign selection depending on past patterns, this chat would be like correctly speaking Chinese without understanding it. Searle’s (1980) Chinese Room argument puts into doubt structural analysis by computerised artificial intelligence: availability of a whole set of Chinese symbols (a word space) together with a code for manipulating these symbols (the algorithm; syntax) may predict the correct response to a sign as stimulus without understanding it. Formal sign systems help us to use systematic properties of the text but this does not yet amount to understanding its meaning. Understanding requires dual symbol grounding: anchoring the symbols directly into their referents (semantic) and into human purposes (pragmatic). This anchoring depends not only on common sense – the inter-subjective connections made by the senses of other interpreters like ourselves – but also has to be a sensorimotor and moving phenomenological experience to avoid infinite regression of inter-

referring signs (Harnad 2005). Texts are conventional expressions of the lived experiences of the author. They tell both about the social context of their production, and provide us with the means to share experiences.

Interpretation

Hence, the connection between the signifier and its signified is both denotative, referring to 'literal', 'obvious' or 'common-sense' meaning, and connotative, referring to figurative, socio-cultural, and emotional associations (Barthes, 1967). Smaller structural features must be grounded to the examination of larger features of the text, such as narrative, rhetoric or ideological discourse. These higher orders of text are often the key to interpretive TA in the social sciences. **Narrative** categories such as actors, actions, events, contexts and the moral of the story allow us to see through the workings of particular stories, and see the commonality underlying a variety of stories from very different contexts. **Rhetoric** offers a different set of categories such as inventions of argument (logos, ethos and pathos), particular genres, composition and tropes. Here the function of public persuasion of texts comes to light. Finally, the analysis of **ideological discourse** offers yet another set of categories which reveal how reality is selectively framed, subjects and objects are positioned, and issues are masked, silenced and written out of the picture. The connotative nature of textual analysis necessitates the interpretive element for understanding the meaning behind the structure.

Addressing the dichotomy between structural analysis and interpretation, Eco (1992:63) suggests that the analyst follow an abductive logic including the triple intentions of the text, the reader and the author. Abduction is a process of hypothesis building from insights and clues of structural patterns. This process engulfs the analyst-reader in a dialogue with the text and the author. Thus, meaning is `forthcoming` from activity rather than being `discovered`. As text analysis is an exercise, understanding a text requires both explicit operations and implicit intuition. Rather than a one-shot hypothetico-deductive prediction, abductive inference requires the meticulous examination of different structural patterns in the text. The logic of ‘abduction’ resolves the tensions between interpretive and structural analyses and offers a re-formulation of the "old, and still valid hermeneutic circle" (ibidem, 1992:64).

Pierce introduced ‘abduction’ to chart a third way of logical inference after deduction, deriving valid conclusions from certain premises, and induction, inferring general rules from

observing particular cases. Abduction seeks no algorithm, but is a heuristic for luckily finding new things and creating insights. Interpretation as abduction defines that logic of insight. Abduction infers from observed results to an observed case on the basis of an ad-hoc invented rule, i.e. the interpretation. We find that the rule is consistent with the patterns; we then conclude that the rule explains the patterns, having discarded some alternatives. It is also known as inference to the most plausible explanation (Harman, 1965). Abductive logic does not replace deduction and induction but iteratively bridges them. Interpretation involves both practical and playful activity; work in conjunction with play solves puzzles (see Lenk, 1993).

Hence, the second dimension of TA reminds us of the allocation of time and resources; there is a trade-off between securing the structural features and jumping to plausible but uncertain conclusions. Abduction teaches us not to reach conclusions before we have secured enough structural features. This leaves open the question of which and how many features to secure. But it also points out that we have to dare best insights under time pressure, however hypothetical that might be; time is short and full structural analysis can take a very long time (‘ars longa, vita breve’). We must at times dare a conclusion on the available evidence. Here, the benchmark is the power of our interpretation to enlighten, persuade, or inspire the audience in a particular situation which is often supported by the visualisation of our text data.

Quantitative and qualitative

Faced with a plethora of approaches to textual analysis, researchers can be tempted into considering these as falling onto one side or another of a divide between the quantitative and the qualitative. This distinction is superficial and is perpetuated due to two interrelated factors: first, the general aim of eliminating ambiguities in research has led to an over-zealous effort to categorise methods as well. While such efforts can be worthwhile for didactic purposes, the concerns raised may sometimes be counterproductive for actual research. Second, one might argue that the distinction continues simply as a de facto convention, born of various traditions in positivist and interpretative research. However, these are epistemological reifications that burden the deliberation of methods. They come from confounding data collection and analysis with principles of research design and knowledge interests. A positivist can pursue qualitative data collection and analysis such as focus groups

while an interpretivist can ground his/her analysis in statistical tools such as analysis of cross-tabulations, clusters and similarity measures.

In academic practice, ambiguity is something to be avoided at all costs. This approach certainly has its value and uses; perhaps, for pedagogical reasons. Course syllabi, for instance, very precisely demarcate numerous, highly-specific qualitative and quantitative skills and techniques for dealing with various types of data and research questions. A student faced with a problem is thereby expected to resolve it through simply knowing and applying the right quantitative or qualitative technique; and the expectation is that this will work, like magic. With textual analysis, however, it is often forgotten that the qualitative/quantitative distinction is motivated by the misconception that examining meanings can or ought to be completely different from examining words.

Added to this, it appears easy to simply associate inductive and interpretative works with ‘qualitative’ research and hypothetical-deductive, statistically-based analyses with ‘quantitative’ research. Certainly, many scholars in the field of textual analysis have little hesitation in branding their work as essentially qualitative or quantitative (see Mayring, 2000, or Schrieier, 2012).

This is ironic because many authors after highlighting (often in introductory chapters of books and articles) the futility of categorising analytic works as either qualitative or quantitative, implicitly build the inductive-deductive dichotomy into their argument. So a key problem is that researchers in mixed-methods research do not then actively promote convergences in their work, but consider their work as qualitative if working with ‘soft’ textual data.

An unfortunate consequence of juxtaposing qualitative and quantitative paradigms has been the uncanny emergence, escalation and entrenchment of a contest between self-proclaimed methodological camps. Branding in this vein is typically used to claim the putative superiority of one approach over the other. What is potentially dangerous is that such efforts are driven by a misconception that the two approaches are intrinsically incompatible, which sees many scholars self-identifying as either a ‘qualitative researcher’ or ‘quantitative researcher’.

A pervasive view in this artificial contest between the qualitative and the quantitative, for instance, is that one ought to consider the interpretative process juxtaposed against the

process of rationalisation. Interpretation is thereby often associated with such things as creativity and imagination, while rationalisation is equated with logic and numerical categorisation. However, even when words are not transformed into numbers, interpretation still proceeds along the same lines of rationalisation, including systematic readings, transparency and methodical reportage when a text is analysed. Logical thought and rationalisation are not exclusive to mathematics or numerical data-handling, but are a crucial part of the interpretative process as well.

Hence the purported dichotomy between qualitative and quantitative is spurious because, firstly, no quantification is possible without a priori qualification and, secondly, no quantifiable explanation is possible without a posteriori qualitative analysis. From the outset of any research process in the social sciences, one requires a notion of qualitative distinctions between social (or, in textual analysis, semantic) categories before one can measure how many words belong to one or another category. Similarly, in the final and perhaps crucial stage of any analysis, it is the interpretation of outputs that is the key to making sense out of it all – and here, the more complex a statistical model, the more difficult the interpretation of the results (see Bauer, Gaskell & Allum, 2000).

5 A possible demarcation between Content Analysis and Text Analysis

As the former entails the latter, the relation between TA and CA can be seen in various ways. We could consider these terms as hierarchical, the one containing the other. In this sense, CA is simply a specific form of text analysis. Alternatively, CA and TA may be considered as different ways of dealing with textual material. The difference might arise on a number of dimensions, such as quantification of content, formalisation of procedures and logic of interpretation and in the role attributed to the researcher.

In table 1 we offer a four-fold classification of procedures for dealing with texts. We differentiate two dimensions: the qualitative and quantitative axis in the horizontal, and the content analysis (CA) and text analysis (TA) axis in the vertical direction.

a) We identify CA with a focus on denotative meanings: words denote concepts. Its focus is semantic, and the logic is deductive, i.e. it works from a pre-established

coding system derived from a conceptual framework. It assumes that text refers to an external reality. Textual production puts meaning into text and CA takes it out again. CA starts with a predefined framework and is therefore 'etic': an outsider-looking-in point of view (see for example: General Inquirer[2]). CA is best characterised as top-down way of seeking information guided by predefined conceptual framework. It emulates the hypothetico-deductive logic of survey research from respondents to text units.

b) By contrast TA focuses on connotative meaning, the circulation of symbols, and follows an inductive or abductive logic. Its perspective is 'emic'; tries to understand intentions of the author, the text itself, and of the reader/audience from their perspectives. TA is a more bottom up, heuristic analysis, supporting an interpretative process rather than revealing 'facts' of the text. TA is more concerned with the symbolic than the conceptual meaning of texts. Texts are cultural artefacts that actively construct 'actuality' as distinct from 'reality' by using symbol systems. Hence, TA focuses on relational and pragmatic aspects of texts rather than their content. Its focus is on co-text and context, linking the elements to larger units (themes to paragraphs, paragraphs to texts, texts to text corpus, text corpus to social contexts etc...). Abductive logic iteratively interprets how these layers interact with each other. It brings together the structural logic of semiotics with the interpretive logic of the hermeneutic circle.

[2] http://www.wjh.harvard.edu/~inquirer/

Table 1: how to distinguish CA from TA

Textual Analysis **Methodologies**	***a) Content Analysis*** Denotations, concepts Etic, focus on purpose of the research Top-down categorisation Hypothetico-deductive modelling	***b) Text Analysis*** Connotations, symbolism Emic, focus on understanding Bottom-up categorisation Abductive modelling
Quantitative, numerical Statistical or graph-theoretical formalism	**A1:** Hypothetico-deductive modelling Dictionary based analysis Relational analysis of narratives Prediction	**B1 :** Abductive modelling Word-space model Corpus linguistics Text-mining Automatic pattern detection
Qualitative, non-numerical informal	**A2:** Thematic analysis with predefined index system	**B2:** Hermeneutic reading Interpretation Grounded theory Open indexing

A1: **Quantitative CA** operates from a pre-established coding frame; the coding process is closed; after a period of piloting, no additional codes are allowed in the coding process. This includes mechanised procedures such as assigning keywords to categorisation dictionaries as in General Inquirer or similar KWOC (keyword out of the context) type analysis. Many categories in these analyses represent grand theory concepts such as ‘modernisation’ or ‘values’ which were prominent at the time when CA was developed by the founders of the method such as Lasswell, Bales, Berelson, Gottschalk, Festinger, and Osgood.

Our collection will not cover this material, because it is the focus of Franzosi’s (2008) collection on CA. Franzosi’s volumes underline the quantitative aspects of classical CA as “a technique of measurement applied to text” (Markoff et al., 1975: 20, 35–38). The early

canons of CA methodologically emulate scale development in their efforts for building coding schemes to operationalise abstract theoretical concepts “to arrive at rather unambiguous descriptions of fundamental features of society” (Lasswell, 1941: 1, 12). In many respect, CA adapts survey data collection methodology to text analysis, creating a matrix of sampling units and variable values. It follows similar sampling and measurement techniques. As for analysis, CA applies statistical hypothesis testing to make “replicable and valid inferences from data to their context” (Krippendorff, 1980: 21). In this view, CA aims for a ‘scientific’ approach; the analyst is an expert intending to reveal factual reality behind words. The purpose of CA is to predict the beliefs, desires and intentions of the text producer or the underlying social phenomena rather than interpreting the text. Words are just symptoms for an underlying latent structure. This approach undervalues the interpretive element as it aims to reduce meaning to denotation. Franzosi (2004: 231) highlights this dilemma and calls for relaxing the hypothetico-deductive logic of classical content analysis. He emphasises that quantitative text analysis should concentrate on bringing out novel patterns in the data rather than ritually sticking to hypothesis testing. He also points to the potentials of rhetoric and frame analysis which are basically interpretive methods in the construction of CA coding schemes (Franzosi, 2008: xxxv). This potential which he states as a future prospect approximates his approach to ours.

A2: **Qualitative content analysis** envisages a coding process where the categorisation system is pre-established, but only in part; building the coding frame is relatively open. We might call this for the moment ‘thematic coding with a preliminary index system’ (operating like a library classification catalogue). This is often used in the coding of interview transcription or streams of documents with a determined theoretical outlook in the research. Qualitative content analysis can be considered as an extension of quantitative content analysis where the machine coding falls short. It aims to complement the systematic nature of the former with the qualitative-interpretative steps of analysis by replacing the rigidity of the machine with the resilience of human coders (Mayring, 2000).

B1: **Quantitative TA** focuses on inductive generation of categories or clusters of words with automatic pattern detection techniques for getting the structure of the text from inside the material. Examples of this are semiometry, lexicography, corpus linguistics and semantic-space models with programmes such as ALCESTE. We would also locate the burgeoning field of text mining in this quadrant.

B2: **Qualitative TA** covers the traditional territory of semiotics and hermeneutic operations. The focus is on understanding the intentions of the text, its author and its audience, including the analyst, from their own perspectives. This might comprise deconstruction and grounded theory that aim to transgress the boundaries of any preconceptions during the interpretation process. However, the possibility of understanding without preconceptions is a controversial issue in text interpretation as we have discussed. Without going more into details, it suffices to say that these debates are more philosophical than methodological and the collections in the "Benchmark" section thoroughly discuss this issue. In qualitative text analysis, each philosophical position creates its own methodology, so the analyst can choose or even create the one which best suits to his/her own creed. For example, as Hoggart, Lees and Davies (2002: 165) note for discourse analysis, qualitative text analysis is "something like bike riding…which is not easy to render or describe in an explicit manner".

7 Programmatic elements to think about for text analysis

Through teaching of TA for several years, in discussions over collating this bibliography of key texts, and through several rounds of the annual LSE TMM (Text mining meetings) with researchers and tool makers, we came to the conclusion that TA needs a programmatic statement to cope with the proliferation of activities, materials and procedures from many different disciplinary corners of the social sciences and increasingly from Artificial Intelligence and ICT experts as well.

Our manifesto for TA is built on four points that deserve attention:

1. Clarification of TA language

Text is not a material for which the social sciences has a monopoly of competence. On the contrary, it is a material that is widely shared across many different disciplines including linguistics, the humanities, social sciences, and increasingly information technologies and artificial intelligence. This creates a proliferation of terms and concepts that confuses the researcher and certainly the student. To avoid undue 'tribalism' forming around particular terminology, we caution against building social identities for example over the use of words

such as ‘tagging’ or ‘coding’ when labelling texts, unless we have gained a good understanding of whether these distinctions are crucial. If the distinction is not crucial, then we should create a dictionary of synonyms and focus on the real distinctions. We expect that a clarification of key terms across all disciplines dealing with texts will help along the efforts of TA.

2. Clarifying the role of the human element

In the social sciences the role of the coder has always been slightly precarious. Through the concern for reliability, the human interpreter/coder has been seen as an inevitable evil, a source of error to be replaced by machines one day. The confidence in this solution arises from a measurement perspective, such as psychometrics, which axiomatically declares that the level of reliability defines the upper limit of the level of validity of a measure. Under this logic, the first step to increase validity is to maximise reliability by automating and standardising the human coder to the maximum. No human, therefore no measurement error! The computer has been hailed as the solution to this problem with the hope that algorithms could replace the unreliability of a moody, tired or untrained coder. This utopia of a perfect reliability again raises its head in the current enthusiasm for machine reading, information retrieval systems, text mining and computational linguistics.

In this context, we need to reflect again, as others have done before (Markoff, Shapiro & Weitman, 1974) on the indispensability of the human mind for understanding. TA includes both feature detection and understanding, making up one’s mind and drawing conclusions that amount to an interpretation. Human understanding is necessarily abductive and hermeneutical when making sense of symbols. The more we know, the more we are immersed in the text, the more it signifies. Moreover, reading is an embodied experience; we put our understanding into practice, associate our phenomenological experiences with those which the text arouses. Understanding requires an active dialogue between the text and the reader. Although machine reading and coding can pass the Turing test, recognise a set of symbols and assign a symbol (code) to it according to rules as good as humans do or better, it can hardly pass Searle’s (1980) Chinese Room test; it can only “chat” with the text but cannot enter into a dialogue with it since this requires the making of sense.

However, reading a text should not be considered as merely a sensual activity and an endless deconstructive playground between the reader and the text. Imposing imaginative associations upon a text will end up in an ‘infinite interpretive drift’. The interpretation process should be limited to the ‘internal textual coherence’; the integrity of the text should be a benchmark for the interpretation of other parts of the text. This brings forth structural analysis where the machines and automatic pattern detection techniques can contribute. This makes text analysis an abductive process involving a trialogue between human, machine and text.

3. Foregrounding the abductive logic of TA

A corollary of the former point asserting the human-machine-text trialogue during TA is our focus on foregrounding the logic of abduction. Much social science methodology operates on a language that seems to force a choice between deductive and inductive methodologies. We reject this language as one of forcing a false choice, and operating with the fallacy of the excluded third. TA does not face a dilemma between the Scylla of deduction on the one hand, and Charybdis of induction on the other. We suggest abductive logic as the middle way out of this forced choice: the logic of inference to the most plausible explanation of the given evidence, considering less plausible alternatives. As it entails both machine inference and human intuition, it can maintain the human-machine-text trialogue.

4. Operationalising higher-order concepts such as framing, metaphor, narrative, argumentation and discourse

Our ambition remains also to rescue the intuitions of the significance of higher order concepts of TA, such as framing, metaphors, argumentation, rhetorical proofs and ideological discourse, integrating them into the age of computerised TA. As routines for computer assisted TA proliferate, we have to be careful not to get caught in the law of instrument or the functional dependency of thinking: letting the tool determine what we can think about. If text analysis is defined by the available computer algorithms, we might well fall into the trap of the young boy who knows how to handle a hammer, so everything he comes across appears to be in need of hitting.

Keeping up the quest for analysis of higher order concepts has a dual function. On one hand it reminds us of the aspirations of text analysis in the social sciences, to recognise the functions of framing, narration, rhetoric and ideological-deluding discourse in written materials. On the other hand, it offers guidelines on where the software and method development has not yet reached. It defines the objectives for method and tool developers on where to go from here. We can now access easily the association structure of a text through co-occurrence analysis of the vocabulary, but we do not know how this extends to the narrative structure of this text. The aspiration of higher order text concepts defines the frontiers and creative tension for tool and method development.

These elements of a programme and a 4-point manifesto for TA in the social sciences are the outcome of our combined and collective search for key texts to which the aspiring social scientist should have easy access.

8 The Order of the Text Collection

To reach our collection of texts we went through several rounds of collecting, discussing, classifying, reducing, expanding again, and querying the selection, matching it to an emergent conception of the field. We ended up with the following logic of classification that matches our reasoning as summarised in table 2 below:

Table 2: The six section of the collection of key papers

1 Foundations	2 Text Preparation	3 Approaches	4 Mark-up logics	5 Applications	6 Validation
cultural indicator benchmark issues	corpus construction	word space models narrative rhetoric discourse	content coding thematic indexing	political science sociology & psychology economics & marketing mass media studies	similarity triangulation abductive logic

1. **Foundations:** This part of the collection provides the texts contrasting essential approaches to text analysis: reading and using. We include texts that give more in-depth insight about the controversies about these issues.

Benchmarks

The first part of the collection focuses on the fundamental texts discussing the controversies concerning the reading process. They question whether it is possible to formulate general rules for discovering the “true” meaning of a text. Is there a scientific method for securing some kind of objectivity when analysing texts? Is it possible to arrive at a “correct reading" of a text ruling out any other rival readings? The “Verstehen” (i.e. German for understanding) approach, a benchmark to distinguish social from physical sciences, gives a negative response to these questions: observation of an act is not enough to fully infer its meaning. Reading an act requires the comprehension of the mind sets of its producers and comparing theirs to ours. In our collection, Theodore Abel discusses the vagueness of the “Verstehen” concept besides its wide usage to distinguish social sciences from physical sciences. He concludes that although the operation of Verstehen performs some auxiliary functions in analysis, it lacks the fundamental attributes of the scientific method. Therefore, it does not provide new knowledge and it cannot be used as a means of validation of an inference. For Umberto Eco, on the other hand, not an objective, but a systematic way of performing Verstehen is possible. During this performance, understanding the mind-set of the audience for whom the text is produced is equally important as the author’s. The act of reading is not a passive transfer of meaning but occurs through a dialogue between reader and text. A text is not produced as a fully cohesive connection of propositions but made of sparsely connected meaning units. Despite the many gaps within the texture of meaning units, texts need to be coherent to make sense to an audience. The author writes the text for a Model Reader who is coherently able to decode the missing links according to their cognitive capacities, lived experiences and cultural conventions. Hence, understanding the meaning of a text requires comprehension of both the author’s and the Model Reader’s mind sets. According to Hans-Georg Gadamer, objective understanding is never possible: interpretation is based on the implicit mind set which is reflected upon the text by the person reading it. The meaning of a text changes as the historical consciousness, the mind sets determined by the socio-historical context, changes. The interpretation process is a fusion of horizons where the analyst finds the ways to compare the historically effected consciousness of his/her time with the one when the text was produced. Understanding is neither a subjective nor an objective act but a process where the past and present mind sets bounding the meaning of a text of are constantly negotiated. According to Paul Ricoeur, complete analysis of our preconceptions is an impossible task since there is no unmediated self-understanding which also is the subject of interpretation. On the other hand, we need a sense about the whole of the text to understand the part. Since the whole is never fully complete, we start with an educated guess about the meaning of a part

and check it against the whole and vice versa. All interpretative activity is then a dialectic process of guessing and validating. Hence, there may be conflict of interpretations made even by the same person. Hirsch's contribution to this controversy is the distinction between criticism, an evaluative act determined by the value judgements, and interpretation, which is the relevance of the reconstruction of the author's intention. While the former is subjective, the latter can be objectively established by applying certain normative principles to the understanding process. These principles can be accomplished by determining how the intention of the author is reflected upon the text, and revealing the genre, a sense of the whole, and typical meaning-components, which the work belongs to. Wimsatt and Beardsley point to two important fallacies which we can commit when interpreting a text. The first, intentional fallacy, reduces the text to its conditions of production. It begins by trying to derive the standard of criticism from the psychological conditions of the author and ends in biography and relativism. The second, affective fallacy, reduces the text to the effects it evokes on the audience. Both fallacies often produce sweeping arguments about the text itself, and end up with interpreting a text by introducing one's own understanding into and onto the text. Over-interpreting a text, reflecting what one hopes or feels it should say is called eisegesis, and Wright contrasts this to exegesis, what a text actually says. Finally, Skinner discusses what is meant by the process of "interpretation;" why it is necessary to undertake this process at all and whether it is possible to lay down any general rules about this process. He argues that interpreting the meaning of a text requires taking into account factors other than the text itself and discusses what should be the factors that need to be taken into account. However, he also considers the text as an autonomous object linked to its producer who has an intention in mind during the production process. The interpreter needs to focus on the writer's mental world, the world of his empirical beliefs.

Cultural indicators

The texts in the cultural indicators section explore how textual material can be used to extract indicators reflecting the context of their times and cultures and what might be at stake when reducing the meaning to quantitative indicators. For some time, the social sciences have mobilised written materials to examine modern culture for the purposes of mapping variations across temporal and spatial contexts. One is reminded of Max Weber's (1911) old advice to the culturally interested social scientist: take your scissors and start cutting up

newspapers.[3] In this part, Bauer discusses how the systematic analysis of intensity and the contents of the media coverage of an issue over time may help to complement public opinion surveys. Similarly, Beniger draws attention to the importance of the media in public agenda setting. Analysis of media content can give us important clues about public attitudes and opinions and help us to produce indicators of social change. Gerbner and Klingeman et al. emphasise how text analysis can compensate for the lack of other data allowing the examination of the long running cultural trends; Janowitz insightfully predicts today's big data environment and highlights that interest in large-scale and continuous monitoring creates new needs that survey research cannot meet. He highlights the potential of content analysis for the policy making process.

2. Text preparation

Corpus construction

The third section deals with prosaic matters of text preparation before the analysis can begin. Atkins et al. offer an in-depth study of corpus design criteria by picking out the principal decision points, and to describe the options open to the corpus-builder at each of these points. Althaus et al. point to the inherent difficulty in random sampling of text content and draw attention to the significance of news indexes as critical research tools for tracking news content. Besides their usefulness, researchers who use indexes to collect their documents are limited by the categorisation made by the index writers. Althaus et al. test the reliability and validity of the New York Times Index, in locating the relevant text content, and how consistent are the subject headings and index entries as proxies for the full text. Bauer and Aarts argue that statistical random sampling would be inadequate for qualitative data collection that is mostly concerned with varieties in belief systems and social practices; for such incidents it would be difficult if not impossible to define a population and sampling frame in advance. Random sampling requires assumptions about the distribution of already known attributes, while qualitative research seeks to determine these attributes in the first place; the purpose is not distribution of attributes, but their rich characterisation. Corpus construction thus replaces random sampling as the systematic data collection methods for qualitative researchers. Barthes' text is a classical handling of the issue of corpus construction for semiotic analysis. Bieber also addresses a number of issues related to achieving 'representativeness' in linguistic corpus design. He emphasises the priority of theoretical

[3] See Krippendorff (2004), on page 4, referring to the Max Weber's address at the first meeting of the new German Sociological Society in Frankfurt.

research in corpus design which should be complemented by empirical investigations of variation in a pilot corpus of texts. Corpus construction proceeds in circles going in between data collection and empirical investigations. Finally, Valsiner argues that the issue of representativeness of qualitative data remains problematic. Errors in representation can be diminished by the correction of the methods by direct experiential access to data, guided by the researcher's intuition. Any data ultimately is a 'representation of reality' and needs to be treated as such, not only by truth value but also by its pragmatic use value. This implies that corpus construction cannot merely be a linguistic effort but also requires the involvement of the language users in the corpus construction process.

3. Approaches to Text Analysis

Although text analysis should be a systematic effort, there is more than one way of exercising it. The way the analysis proceeds reveals its epistemological and methodological perspectives. The third section gathers four broad approaches to consider text: the word space model, narrative, rhetoric and discourse. While the first two are more convenient for a structuralist perspective, the last two are more convenient for an interpretive perspective. However, these are not mutually exclusive categories. Triangulation and abduction in text analysis (see below on validation) might involve several of these approaches during different phases of the analysis. Each of these higher text notions has developed into a text analysis framework with overlapping 'language games'. This is a key section of our collection. We invite readers to appreciate the approaches and perspectives that are on offer as ways of 'framing the text as X'.

The word space model

This approach offers statistical analysis of vocabularies and semantic networks arising from spatial associations of words, and shows how text can be classified on the basis of elemental or structural similarities. On the problem of what is similarity, we return with validation. Much of this goes under 'text mining' in current jargon. These spatial models are supported by statistical procedures of clustering and factoring (Lahlou), deal with textual features in quantified and numerical forms (Roberts), and as such they can be processed with mathematical formalisms such as network logic (Popping, Diesner & Carley) and become amenable to visualisation.

Narrative analysis

This approach focuses on the ways in which people represent themselves and their worlds to position themselves in the social space and to construct identity. Since narratives are social constructions, they give us important clues about the context of specific social, historical and cultural locations of their producers. Narrative analysis can be both structural and interpretive. According to Propp narratives are structured and they can take different forms. The Fairytale is the archetypical form which is central to all story telling. The structure of the Fairytale is not determined according to the type of the characters or events but by their functions in the plot that can be handled in few categories. Labov and Waletzky also follow a structuralist approach but focus on the story grammar. They combine grammatical elements with sociological features. For example, type of the clause usually gives us important clues about the narrated social positions. Riceour carries the narrative approach over to interpretation. People use narratives to say something to others bounded by structural features. A narrative always involves an author and an audience as well as a statement about reality. Therefore, narrative analysis requires both the objective analysis, for which structuralism provides a tool, and an interpretive element. For Ricoeur, even the presentation of the historical facts themselves are 'fictive' and therefore subject to the reconstruction through imagination and interpretation. Schlegoff gives us a detailed account of narrative research, and worries about its de-contextualisation in structural analysis, and Laszlo applies narrative analysis to psychological research.

Rhetorical analysis

This approach employs the principles of rhetoric to examine the interactions between a text, an author, and an audience. The papers on rhetoric clarify the language game of an old pursuit dating back to the classical period of Ancient Greece (Barthes, Bitzer). A key dimension has been the 'logos', the types of argumentation that are convincing but still formally distinct from deductive or inductive logic (Toulmin); useful analytic advice arises from this practice (Simosi) and also for the analysis of metaphors (Lakoff). A recent revival of rhetorical topics is 'frame analysis.' Frame analysis brings a number of related but sometimes partially incompatible methods for the analysis of discourses (Scheufele). Frame analysis aims to extract the basic cognitive structures which guide the perception and representation of reality underlining a text. Frames are usually latent structures that are not

directly perceivable by an audience. Therefore, framing is more a tacit activity than a deliberate effort. When we frame, we do so by tacitly selecting some aspects of a perceived reality and making them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation (Entman).

Discourse analysis

This approach has recently become a popular focus of research in many disciplines of the social sciences. Fall in quality and malpractice usually follow popularisation and we can see that the terms 'discourse' and 'discourse analysis' have come to be used and abused in widely arbitrary and divergent ways. Biber calls for a systematic approach to discourse analysis and to merge it with the analytical goals and methods of corpus linguistics for the purposes of identifying the general patterns of discourse organisation that are used to construct texts. Potter & Wetherell offer methodological steps for practising discourse analysis. Their text is more focused on the textual analysis concerning the discursive construction of reality. Critical discourse analysis (CDA) on the other hand focuses on how social power is abused, dominance and inequality are enacted, reproduced, and resisted by text and talk in the social and political context. Van Dijk's text is a general presentation of the essentials of CDA, Fairclough lays down an analytical framework for CDA and Hajer offers an application of CDA to policy research with methodological benchmarks.

4. Mark-up logics

The fourth section deals with what we might call techniques to mark-up similar parts of a text for further searching, comparison and analysis. We have identified two major traditions of text-to-code transformations or text tagging: content coding and thematic indexing. Here much confusion over vocabulary and terminology arises from disparate attempts to demarcate for good and not so good reasons different operations and procedures. Although these terms are frequently used interchangeably, we make a tentative distinction for operational reasons. We name the operations for labelling text segments with identical meanings according to a predefined categorisation system to produce some descriptive indicators for the purposes of counting and comparison as coding; and operations for cataloguing text segments so that they could easily be queried, retrieved, sorted, reviewed, or prioritised for further reading as

indexing. Briefly, coding is assigning text segments to classes and indexing is assigning themes to text segments.

Coding

Content analytic coding is said to be 'deductive', i.e. deriving its content coding categories from outside the text. It codes the text units to explicit rules of 'one text unit – one code value' into a data matrix which researchers recognise from survey research, the cases by variables matrix. Here we gather papers on the conceptual basis of CA (Krippendorff, Markoff), that exemplify the ambition of measuring the evaluative attitude and positioning of texts (Janis) on the basis of pre-defined and thus closed coding system of categories.

Indexing

Thematic analysis (TA) is said to be 'inductive' and inspired by grounded theory that is free from any assumptions or pre-conceptions (Charmaz), i.e. deriving its index system bottom-up. Thematic indexing has an open- bottom up ethos that is often pitched against content coding with its closed-top down coding system. But in reality TA with its operational hierarchy of basic, organisational and conceptual codes drifts somewhere between these polemical poles (Hsieh & Shannon). TA shows concerns for 'issue salience' (Buetow). Attride-Sterling offers an important analytical tool for presenting indexing systems as thematic networks: web-like illustrations that link the main themes that constitute a text.

5. Applications and examples

The fifth section brings together examples of applications of textual analysis from different fields of enquiry. We have chosen papers that illustrate analyses of the larger text intuitions of section three. We asked: how does each field of enquiry apply the approaches discussed in that introduction? Political science is concerned with news coverage and debates, actor positioning and issue framing on issues such as nuclear power. For sociology and (social) psychology we cover studies of science news, suicide notes and poverty. The world of economics, business and marketing is keener than ever on text analysis. Here we gather studies on material values, emotions at work, and of mental models. Mass media research is illustrated by analyses of metaphors in Roman texts, and in relation to stem cell research, genetically modified organisms and climate change.

6. Validating the Results

Our sixth and final section deals with the validation of the analysis. In our perspective this is a wide open issue. We do not as yet command clear and defined procedures, if there ever will be. We consider validation a matter of due process rather than an achieved correspondence between model and data, or a fit between model and reality. The issue is thus less one of 'validity' and rather one of 'validation' of text analysis. Our texts raise issues and define the problem along three lines: similarity, abduction and triangulation.

Similarity

Much text analysis hinges on a judgement of similarity between meaning units. Ultimately, text analysis is a categorisation process for recognising, demarcating and understanding these units. Categorisation is based on similarity and dissimilarity. However, the notion of 'similarity' needs clarification (Tversky, 1977). Wallach distinguishes between potential and psychological similarity. The former judges the similarity of two objects or events in terms of the number of common attributes they are found to display. The latter is a more complex process which selectively handles the complexity of the environmentally available attributes with some cognitive heuristics. Depending on experimental research he defines psychological similarity in terms of perceptual assignment to a common category rather than evaluating each of the attributes. Similarly, from a different angle, Eco argues that categorisation is conjecture about the attributes of a series of apparently disconnected elements. Assigning a text element to a category involves reconstructing it in terms of "fair guesses" about lost sentences or words. This argument has important validity implications for text analysis: to categorise a text unit we may either use an already coded rule to which the unit is correlated by inference (the hypothetico-deductive way) or we can provisionally entertain an explanatory comprehension from a text unit to infer rules for categorisation which has to pass further testing (the abductive way).

Abduction

The logic of iterative abduction would be the most appropriate explication of what is involved in interpreting texts on a hermeneutic cycle. In the account of Eco & Sebeok (1983), abduction describes the way the detective orders his or her clues to find the culprit; it is the logic of Sherlock Holmes. Harman contrasts inference to the best explanation (abduction) to enumerative induction which means inferring a relation by simply considering the frequency

of co-occurrences between two events. Establishing a link by only looking at co-occurrences is fallacious since it disguises the fact that our inference is based on certain lemmas, i.e. word units often word stems before grammatical form changes, linking these two events. This statement is an important criticism of the word space approach. On the other hand, inference to the best explanation exposes these lemmas, which play an essential role in the analysis of knowledge based on inference. Hence, abduction compensates the inadequacies of deductive and inductive inference for assigning cases to categories. Kapitan discusses what makes abduction an autonomous mode of inference; it is not based on logical but on pragmatic grounds. Scientific inquiry does not only involve establishing theoretical relations among propositions but also concerns itself with procedures for evaluating inferences to practical directives. The validity of a method can be tested if it can establish relations from which we can infer a question or recommendation that can be legitimate or appropriate for a community of users. Hence, abductive inference establishes relations in terms of descriptions and explanations grounded in everyday practices.

Triangulation

A similar logic of employing different perspectives for cross-checking the soundness of an inference has flourished on methodological grounds. According to classical definition, validity of an inference entails its degree of correspondence to the real world. However, triangulation, rather than testing the truth value of an inference, cross-checks if it can survive the confrontation with a series of complementary methods of testing. Triangulation approaches the same phenomenon from a multi-hypothesis and multi-method perspective. Erzberger & Prein underlining the complementary nature of qualitative and quantitative methodologies, illustrate the advantages of triangulation, focusing on how relationships can be established between different research results coming from applying different methodological approaches to the same problem. Flick stresses the demarcation between validity and triangulation. The aim of triangulation is not to validate our inferences from different perspectives in an eclectic way but a mutual assessment of different analyses to add breadth and in-depth understanding without artificial objectivation of the subject under study. The meaning of triangulation shifts from confirming results to create alternative, sometimes contradictory explanations from different perspectives. This can best be achieved by employing at least one method for exploring the *structural aspects of a phenomenon* and at least one interpretive method which can allow us to understand what this means *to those involved.* Finally, Gaskell & Bauer (2000) showed how and why triangulation has to become

a canonical procedure to secure quality in qualitative research: it guarantees reflexivity as the researchers have to deal with the contradictions.

9 Beyond the Boundaries of TA

We close with a brief word on what this collection of papers excludes. Potential readers might seek something under the heading “text mining” but not find it here. Our collection excludes most of the developments arising from ‘big data’ such as GOOGLE based Culturonomics which uses the massive databases of millions of digitalised books to create indicators of social change. It could be shown how the cycle of fame, the appearance and disappearance once famous names over time, accelerates over the 20th century (Michel et al, 2011). Equally beyond our present concerns is “sentiment analysis”: attempts to mine social media data and shopping comments data to depict collective mood swings, predict economic cycles, stock markets and the next individual shopping move (see Bollen, Mao & Zeng, 2010). Although these developments look interesting, they are heavily computer science and big data based; entirely remote from reading as a dialogue with the text. These approaches have stepped into the realm of ‘using text’ without any consideration of communicative context. As such they transcend our present purpose of documenting TA as an exploration of social processes.

Equally not included in this collection are listings, overviews, descriptions or comparisons of software tools. Text analysis has recently given rise to many different software tools that assist the securing, storing, marking, coding and indexing, and statistical mining of text materials. Some of these tools are used and referred to in this collection of papers. However, for us it was important to separate the logic of text analysis from the implementation of any of these steps in particular software routines. The software is not the method. The taxonomy and comparative assessment of such tools must be sought in other places, not least as any particular text analysis logic might find implementation in different software products, or several analytic logics are included in a single software platform. The latter would be particularly desirable. Text analytic computer support is still underdeveloped, every procedures creates its own software routine and not seldom its own product brand and user community, and no platform as yet exists which covers all available procedures. TA will

enter a new phase once a platform is available that supports corpus construction, tagging, open and closed coding, dictionary and thesaurus based categorisation, linguistic parsing, word space modelling, rhetorical, narrative and discourse analysis, all implemented as user-friendly pull down menus with parameters to select for each routine. The convenient and integrated worlds of SPSS, SAS etc. for statistical routines remain a model for TA as well, and it seems, considering the level of software activism, we might reach there in a few years' time.

References

Barthes, R (1996; [1971] "From Work to Text" in *Modern Literary Theory*. ed. Philip Rice and Patricia Waugh. New York: Arnold.

Barthes, R (1967) Elements of Semiology, London, Jonathan Cape.

Bauer MW and G Gaskell (2008) Social representations theory: a progressive research programme for Social Psychology, Journal for the Theory of Social Behaviour, 38, 4, 335-354.

Bauer MW, G Gaskell and N Allum (2000) Quantity, Quality and Knowledge Interests: Avoiding Confusions; in Bauer MW, and G Gaskell (eds) Qualitative researching with text, image and sound, London, SAGE, pp3-17.

Bollen J, H Mao and XJ Zeng (2010) Twitter mood predicts stock market, 14 October http://www.ccs.neu.edu/home/amislove/twittermood/

Boyce RWD (2000) Fallacies in interpretating historical and social data, in: Bauer MW and G Gaskell (eds) Qualitative researching with text, image and sound, London, SAGE, pp318-350.

Derrida J (1967) La voix et le phenomene – Introduction au problem du signed an la phenomenology de Husserl, Paris, PUF (German Suhrkamp Edition, 2003)

Eco U (1988) Horn, Hooves, and Insteps – some hypotheses on three types of abduction, in: Eco U and TA Sebeok (eds) The sign of three – Dupin, Holmes and Peirce, Bloomington, Indian University Press, pp198-220.

Eco, U (1990a) Lector in fabula, in: The Role of the Reader: Explorations in the Semiotics of Texts, Bloomington, Indiana University Press.

Eco, U *(*1990b) *The Limits of Interpretation. Bloomington: Indiana University Press.*

Eco, U (1990) Some Paranoid Readings. *Times Education Supplement,* June 29-July 5: 705-6.

Eco U (1992) Over-interpreting texts, in: Interpretation and over-interpretation, Cambridge, CUP, pp45-66.

Eco, U. (1997). 'The Kabbalistic Pansemioticism', in: The Search for the Perfect Language, London: Fontana Press, pp. 25-33.

Flick, U. E. von Kardorff & I. Steinke (2004)What is Qualitative Research? An Introduction to the Field; in U. Flick, E. von Kardorff & I. Steinke (Eds.), A companion to qualitative research (pp. 178-183). London: Sage Publications.

Fransozi, R.(2004) From Words to Numbers: Narrative, Data, and Social Science (Structural Analysis in the Social Sciences Vol 22), Cambridge: CUP

Fransozi, R. (2008) Content Analysis: Objective, Systematic, and Quantitative Description of Content, in Sage Benchmarks in Social Science Research Methods: Content Analysis V. 1, Ed R. Franzosi, London: Sage, (pp xii-xxiii).

Gadamer, HG (2004) Truth and Method, 2nd rev. edition. Trans. J. Weinsheimer and D. G. Marshall. New York: Crossroad

Geertz, C. (1973) Deep Play: Notes on the Balinese Cockfight. *In* The Interpretation of Cultures. pp. 412-453. New York: Basic Books.

Roberto Franzosi, *(2008)* Content Analysis: Objective, Systematic, and Quantitative Description of Content

Harman, G. (1965) The Inference to the Best Explanation. The Philosophical Review 74(1), 88-95

Harnad, S. (2005) To Cognize is to Categorize: Cognition is categorization. In: Lefebvre, C. and Cohen, H., Eds. *Handbook of Categorization*. Elsevier.

Gaskell G and MW Bauer (2000) Towards public accountability: beyond sampling, reliability and validity, in: Bauer MW and G Gaskell (eds) Qualitative researching with text, image and sound, London, SAGE, pp336-350.

Kleist, H.von. (1984) [10/12/1810], Berliner Abendblaetter, in: Sämtliche Werke und Briefe in zwei Bänden, (editor H Sembdner), Bd. 2, p338: Muenchen.

Mohr, J.W. (1998). 'Measuring meaning structures', Annual Review of Sociology, 24: 345-370.

Hoggart, K., Lees, L., and Davies, A. (2002) Researching human geography, Arnold: London

Lenk H (1993) Interpretationskonstrukte – Zur Kritik der interpretatorischen Vernunft, Frankfurt, Suhrkamp.

Lenk, H.: Toward a Systematic Interpretationism. In: Stapleton, T. J. (Ed.): The

Question of Hermeneutics. Dordrecht. 1994, 79-88.

Lotman, Yuri M. (2005) On the semiosphere. (Translated by Wilma Clark) Sign Systems Studies, 33,1.

Markoff J, G Shapiro and SR Weitman (1974) Towards the integration of content analysis and general methodology, in: Heise RD (ed) Sociological Methodology, pp1-58

Mayring, Philipp (2000). Qualitative Content Analysis [28 paragraphs]. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1(2), Art. 20, http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204.

Fielding, N and M Schreier (2001) Introduction: On the Compatibility between Qualitative and Quantitative Research Methods. In: Forum: Qualitative Social Research**,** [S.l.], v. 2, n. 1, feb. 2001. ISSN 1438-5627. Available at: <http://www.qualitative-research.net/index.php/fqs/article/view/965/2106>. Date accessed: 01 Oct. 2013.

Hoggart, K., L Lees and A Davies 2001: Researching human geography. London: Arnold.

Michel JB, YK Shen, AP Aiden et al. (2011) Quantitative analysis of culture using millions of digitized books, Science, 331, 6014, pp176-182 [on-line publication 16 Dec 2010]

Reddy M J (1993) The conduit metaphor: a case of frame conflict in our language about language, in: Metaphor and thought (ed A Ortony), Cambridge, CUP (2nd edition), 164-201

Ricoeur, P (1973) *The Model of the Text: Meaningful Action Considered as a Text"*

New Literary History , Vol. 5, No. 1, pp. 91-117

Ricoeur, P (1981). ‘What is a text? Explanation and understanding’, in: Thompson, J.B. (ed.). Hermeneutics & the Human Sciences. Cambridge: CUP, pp. 145-164.

Searle, J (1980), "Minds, Brains and Programs", *Behavioral and Brain Sciences* 3 (3): 417–457

Segre C and T Kemeny (1988) Introduction to the analysis of the literary text, Bloomington, Indiana University Press.

Sloterdijk P (2013) You must change your life, Cambridge, Polity (original German, Suhrkamp, 2009).

Schreier, M (2012) Qualitative Content Analysis. Sage, Thousand Oaks, CA. Sage, Thousand Oaks, CA

Tversky, A (1977). ‘Features of similarity’, Psychological Review, 84(4): 327-352.

Webb EJ, DT Campbell, RD Schwartz and L Sechrest (1966) Unobtrusive measures: non-reactive research in the social sciences, Chicago, Rand McNally.