Advanced Analyses of Text

Análisis avanzado del texto

Emad Jasem Mohamed
Imad.jasim@uomisan.edu.iq
Misan University, Iraq

ABSTRACT

This study aims to analyze texts through analysis of invisible text, which includes many meanings, expressions and norms that can be accessed by symbolic behavior such as language, literature, and the arts. In recent decades, the rapid increase in the development and availability of resources and computational technology has led to an increase in the analysis of multimedia texts that appear in a range of disciplines in clouded in the above study the prevalence of contemporary digital media forms requires researchers to keep sociocultural developments. At this stage, multimedia studies are required as a kind of experimental text.

Keywords: Analysis of text, language, literature.

RESUMEN

Este estudio tiene como objetivo analizar los textos a través del análisis de texto invisible que incluye muchos significados, expresiones y normas a los que se puede acceder mediante un comportamiento simbólico como el lenguaje, la literatura y las artes. En las últimas décadas, el rápido aumento en el desarrollo y la disponibilidad de recursos y tecnología computacional ha llevado a un aumento en el análisis de textos multimedia que aparecen en una variedad de disciplinas en el estudio anterior, la prevalencia de las formas contemporáneas de medios digitales requiere investigadores que mantengan los desarrollos socioculturales. En esta etapa, se requieren estudios multimedia como una especie de texto experimental.

Palabras clave: Análisis de texto, literatura, lenguaje.

INTRODUCTION

Text analysis comprehends a class of techniques for the social scientific study of *communication*. From one point of view, text analysis (content analysis) can be contrasted with *observation*. Central to it is not the material, the visible, the audible. It is the invisible: the world of meanings, values, norms as these are made accessible by symbolic behavior (language, literature, plastic art, music etc.). The '*content*' in content analysis is the material`s meaning, i.e., the social significance behind the visible or audible symbols in papers, books, film recordings, pictures, paintings or behind human activities in radio and television broadcasts or in films. In his book, *Computer-assisted Text Analysis*, Roel Popping states that, in this case, the focus is especially on the content of the message, by preference in mass communication where the message is addressed to many people.

About text analysis

Originally, text analysis was used to draw conclusions regarding the source of the message. Sources were often a collectivity, like for instance an organization or a government. However, communication is more broadly understood as involving message, channel, audience, as well as source. In fact, these 4 aspects of communication represent the most common contextual variables used in analyses of text and transcripts.

Beyond examining symbolic content within various contexts, text analysis is used in bringing structure to an enormous amount of rather unstructured information. This allows the investigator to make explicit various aspects that might not be noticed by a lay observer. Therefore a better understanding can be gained in certain aspects of societal processes.

On the other hand, the meaning of text analysis remains still illusive. One might ask whether sensoty perception could have the invisible and immaterial as its object. The analysis of content does not focus on behaviour. It focuses on recording and comparing artifacts of attributes produced by human behaviour, both individually and interactively. Its fundamental characteristic is the fact that it is concerned with the communicative act *post hoc*.

The Aim

The present study aims at: Advanced analyses of text, is the term for any text analysis in which variables indicate the occurrence of particular concepts.

Thematic analysis

Thematic text analysis (referred sometimes to as *traditional text analysis*) is the term for any text analysis in which variables indicate the occurrence of particular concepts. Practitioners of thematic text analysis usually reserve the term *theme* for broader classes of concepts. The theme can be considered as subject and as attribute. In this case it is a subject; the analysis is focused on a specific referent (ex.: the president, the US, communism, British foreign policy etc.). Themes as attributes are indicated by their measures of qualification. Thematic text analysis allows the investigator to determine what and how frequently concepts occur in texts. The method is particularly useful when the investigator is interested in the prominence of various concepts in texts, possible reflecting broad cultural shifts.

The data matrix in a thematic text analysis has one row for each randomly sampled block of text and one column for each theme that may occur in these text blocs. Let's say the following sentence comprises a text block:

> *I always drive safe.*

If an investigator is interested in the concept *SELF-REFERENCE*, this sentence contains one occurrence of the concept, namely the word *I*. Investigators, though, will most likely have

numerous concepts they are interested in and a dictionary containing list of corresponding words and phrases that reflect each.

Cells in a thematic data matrix reflect the number of occurrences of a particular concept within a specific block of text. An example of a resulted data matrix is illustrated in Table 1.

Table 1. Data matrix for a thematic text analysis

ID NUMBER	Concept 1	Concept 2	Concept 3
1	0	0	0
2	1	0	1
3	0	2	1
4	0	3	2
5	0	0	0
.	.	.	.
.	.	.	.
.	.	.	.

Source: **Popping R.,** *Computer-assisted Text Analysis,* 27

ID numbers correspond to text blocks. If Concept 1 were SELF-REFERENCE and the first text block were the sentence *I always drive safe*, the *I* in the associated cell would indicate a single occurrence of SELF-REFERENCE in this text block. Once concept occurrences have been encoded, the frequency and their co-occurrences can be investigated.

Text Encoding

If the instrumental approach is followed, the dictionaries are available before the actual coding process starts. As these dictionaries contain all concepts and their corresponding search entries, the coding process is easy to automate. The representational approach, though, doesn`t imply on beforehand a predetermined relation between concepts and search entries. These relations and linkages depend on the context within which a search entry is used. Hence, this approach is hard to automate. This approach allows also not just to detect new search entries for concepts in the texts. New concepts might be derived as well from information in the texts.

Contingency analysis

In thematic text analysis one can report occurrences and co-occurrences of concepts. Occurrences indicate   the prominence of   themes.   When compared across contexts   they can afford inferences about culture`s changing themes for instance, ideas, issues and dilemas or differences between media in representation news content about the same issue. Looking at co-occurrences means looking at associations   among   themes.   This   analysis   is also known as *contingency analysis*.   In   this   type   of analysis   the goal   is   to calculate associations among occurrence measures and to infer what the resulting pattern of association means. Problems may occur if these inferences are abour how concepts are related. Roel Popping invites us therefore to consider the following text block is investigated: *'The man likes detective stories, but his wife prefers love themes'*. The concepts MAN (represented by *the man*) and LOVE THEME (represented by *love themes*) co-occur in this block, but no relation between the 2 is specified. For such inferences relations should have been encoded *a priori*, not via *ad hoc post hoc* looks at the text.

KINDS OF TEXT ANALYSIS

Text analysis – Semantic approaches

In the thematic approach concepts are counted. On other side, in the semantic approach relations among concepts are also encoded. These relations are founded in clauses. The advantage of this approach over the traditional thematic one is the fact that it 'preserves' the complex relations between parts of texts and, hence, preserves much of the narrative taste of the original text.

The move from a thematic to a semantic text analysis expands the type of questions that a researcher can answer. Referring to propaganda techniques in making this point, it is noted the fact that in a thematic analysis a possible research question would be the following: *What themes are mentioned in propagnad that aren`t mentioned in other communication*? Using the semantic approach, the question can be extended to *what syntactic strategies are used by political leaders when their policies fail (/succeed)?* Unlike the former question, the last one asks about concrete relations among concepts used in different social contexts.

Semantically encoding data requires that one fit concepts that occur in a clause into a semantic grammar. The concept of clause is the usual one here, more precise a sentence or part of a sentence that contains an inflected verb, an optional subject and/or object plus all modifiers related to this verb, subject and object. Let`s take the sentence *The lower the prices, the mor the customers*. This sentence is comprised of two clauses, because it contains in the first place two inflected instances of the verb *to be*. Therefore, one determines the number of clauses in a text by counting its implicitly and explicitly inflected verbs. A semantic grammar is a template in to which codes for concepts are fit. For instance, let`s take a four-position semantic grammar for encoding opinion statements appearing in newspaper editorials. The grammar specifies the following 4 semantic components:

1. *Agency* – the initiator of an activity
2. *Position* – the position regarding the agency`s activity
3. *Action* – the activity under consideration
4. *Object* – the target of the activity

Applying this template to the sentence *People don`t listen enough to each other*, would require assigning the concept *people* to the component agency, *ought to* to the position component, *listen* to the action component and *people* to the object component. Other examples are *Unemployed people* (agency) *cannot* (position) *improve* (action) *their standard of living (*object), and *gasoline taxex (*object) must *not* (position) be *raised* (action).

Text analysis- semantic grammar

This semantic grammar has the Subject-Modal-auxiliary-verb-Verb-Object form used in linguistic content analysis. In most of the cases valence information (regarding negation, evaluations and so on) is included under the verb component. For this reasone one sometimes refers to a semantic gramar as having a Subject-Verb-Valence-Object form. By taking texts relational characteristics into account, semantic text analysis improves upon thematic text analysis methods and over comes many of its problems. Based on a thematic text analysis, co-ocurrence of subject and object can be identified, in the semantic text analysis the relation is specified and might be investigated (Yule George 2010).

Semantically encoded data can be used to make inferences about the condition under which texts` authors take specific positions on others` intentions. For instance, when a government`s totalitarian leadership initiates a public relation campaign to 'democratize' its image in the US press, such data could be used to test whether a significant increase had occurred from before to after the campaign in the odds that US news stories describe the regime (subject) attempting (position) to listen to (action) its citizens (object) (Popping R., *Computer-Assisted Text Analysis*, 1, 26-29).

Multimodal text analysis

Multimodal analysis includes the analysis of communication in all its forms, but is particularly concerned with texts which contain the interaction and integration of two or more semiotic resources – or *modes* of communication – in order to achieve the communicative functions of the text. Such resources include aspects of speech like intonation and other vocal characteristics, the semiotic action of other bodily resources like gesture (face, hand and body) and proximics, as well as products of human technology like carving, painting, writing, architecture, image and sound recording and, in more contemporary times, interactive computing resources (ex.: digital media harwares and softwares). Different semiotic resources bring with them their own affordances and constraints, both individually and in combination, as well as analytical chalanges in terms of the natures of the media, the detail and purpose of analysis, and the complexities arising from the integration of semiotic resources across media.

Kay O`Halloran and Bradly Smith show us that two major strategies have emerged, constituting two of the major approaches by which multimodal text analysis can be characterised: the first is exploring theory, using text analysis as both test and illustration of the discussion of general principles; the second is exploring actual texts, working form such analyses towards generalizations. The first approach doesn`t assume and problematises theory. The much of the work is involved with comparing and integrating knowledge and practic from often disparate theoretical traditions. Multimodal studies suits such an approach, when one considers the wide range of disciplinary and theoretical traditions for which multimodality is relevant. In this approach theoretical generalization is often applied to analyses across several different texts and types of text, in order to derive, test and emphasize general principles. The second approach is to pay a very close attention to and work from actual specific texts, where the focus is upon the text, building up detailed descriprion, often as a form of running annotaion and usually adapting and applying an established theoretical and descriptive framework but deriving descriptive generalizations out of such text analysis and modifying theory as a result. In this approach, analytical detail is paramount, and once again, the challenges of access, transcription, analysis and reproduction in publication are all to apparent.

While the difficulties of multimodal text analysis are apparent in most works which explore this field, there has nevertheless been built up over time a corpus of detailed multimodal analysis which has provided the means for testing, exploring and illustrating ideas about how multimodal communication works. It`s important to recognize therefore the fact that, while it`s only in recent decades that extensive multimodal text analysis has announced the emergence of a distinct field of multimodal studies, in fact the studies of multimodal communication and artefacts can of course be found in the long lineage of works within anthropology, archeology, art criticism and history (ex.: painting, sculpture, music, theatre, opera, film etc), computer science, engineering, psychology, and all fields of research engaged with human or non-human communication. The related practical disciplines, like the various forms of art, provide also ample material for multimodal text analysts to draw upon. Moreover, in recent decades, the rapid increase in sophistication and availability of technological (particularly computational) resources and techniques for analysis of multimodal text has no doubt driven the fast increase in multimodal analyses appearing within a range of disciplines.

Exhaustive text analysis

It is clear the fact that both the empiricism of detailed, exhaustive text analysis (coping with challenges this raises) and the ongoing problematisation and exploration of theoretical genralization and abstraction are needed for the development of resources for and practice of multimodal text analysis. Moreover, multimodal text analysis requires the integration of both low and high level analyses: as with studies of intonation, the study of multimodality has been centrally concerned with the material plane, the organization of the physical (including technologically-driven) phenomena by which meaning is created, in their relation to higher level grammatical and semantic

organization of such as abstract, semiotic system and structures. How to relate the also-called the content and expression planes has been a central central challenge for multimodal analysts and this challenge has become more acute as sophisticated instrumental analysis – particularly computational automated techniques for feature extraction – make possible the analysis of primarily low-level features within large corpora of multimodal text. The availability of such techniques is both a great opportunity and difficulty for scholars of multimodal comunication: one major challenge is to trace and make explicit the path from such automatically detected features to the socio-cultural patterns significant to multimodal analysis of text, in ways that draw upon the respective riches of computational and semiotic sciences, as well as the knowledge and practices of other disciplines like mathematics, physics, psychology, ethnography etc. The immediate demands of multimodal texts analysis actually become site for the exploration of synergies between disciplinary and theoretical traditions, as it is to understand actual text through the kaleidoscope of the multitude of relevant perspective.

CHALLENGES AND SOLUTION FOR THE ANALYST

Variety of text and curricula

The variety of work and approaches within this range of disparate fields of sciences and humanities research constitutes both a rich resource and a challenge for the multimodal analyst. One must sort through the complementarities, inconsistencies and redundancies of the different approaches and perspectives, working out which types of analysis suit which researh project – some approaches being more appropriate to certain tasks than others. This has been a key issue with respect, in particular, to the adaptation and application of linguistic theory and description, which has been an important influence upon multimodal studies. Multimodal analysis must include analysis of language where relevant; but in the analysis of semiotic resources other than language, whether involved in intersemiotic relations with language or not within a particular text, the question of how much of linguistics can be adapted for the analysis is still an open question. It`s important to note therefore the difference between the application of general theoretical principles, and the adaption of a specific description: It is the former that drives the development of descriptions of semiotic resources other than language, while the latter (linguistic descriptions) can nonetheless provide useful materials, if appropriately adapted for such descriptions. The same question, though, can be applied to work within all relevant disciplines: determining the affordances and constraints of theories, descriptions and methodology adapted and applied from existing disciplines and traditions is a crucial challenge for multimodal text analysts (Fromkin 2017).

Multimedia text challenges

Other issues remain for the multmodal text analyst. For instance, the issue of corpus constraint keeps challenging multimodal text analysts: gathering multimodal corpora has become increasingly difficult, not the least because of ethical concerns which become acute when audiovisual recording of naturally occurring discourse is available. There are also difficulties in terms of data collection (like the recording process itself, often involving technical expertise not always at hand for multimodal analysts), and the dynamism and hyper-textuality of contemporary interactive digital media has compounded this issue. The issues of detail, scope and complexity continue also to haress multimodal analysts, but becoming increasingly evident with the availability of sophisticated interactive digital software application which provide platforms for an ever-more minute analysis of multimodal phenomena. The human analyst can never be replaced by computer based or other technology based approaches, but such resources increase the power of analytical reach (ex.: via low-level feature extraction, mathematical processing, visualization techniques), presenting richer but more complex and detailed phenomena to analyse.

The application and value of multimodal text analysis are enormous. Multimodal communication is central to human existence, and yet such is the nature of multimodality – the joining of abstraction and material in semiosis – that this is an area of the natural world that still remains relatively underexplred, and not well understood, compared with the material plane.

CONCLUSIONS

The connections between language and social psychology are changing at an accelerating rate. When journals like *Journal of Language and Social Psychology* were founded, most research was based on written text or transcriptions of spoken text, all of which were hand-typed, hand-scored and stored in a filing cabinet for later analyses. Researchers interested in language and social processes have hystorically been trained in laboratory methods whereby participands were run, one at a time, in highly controlled setting to capture best the links between language use, cognitive processing and communication dynamics (Tausczik Y.R., Pennebaker J. W., *The Psychological Meaning of Words: LIWC and Text Analysis Methods*, 38).

Innovations in word analysis are (Dowing & Locke 2006) challeging the social psychological methodologies most of us have grown up with. In the amount of time it takes to run a single participant in a social psyschology language study, now we can download dozens of personal writings, interaction transcripts or other forms of text that can be analyzed in just seconds. The online world provides a way more diverse population from which to draw as well as access to a wide rance of languages.

Research using new text analysis methods will be expanded to capture cultural differences mirrored in language use. Language style conveys subtle information about social relations. The relevant social information can vary greatly between langage and cultures. Of course, some of the most striking cultural differences in language – like markers of politeness, formality or social closeness – are inherent in function words rather than content words.

The proliferation in forms of contemporary interactive digital media and the ubicuity of their use puts demands upon scholars of human communication to keep up with wither socio cultural developments. What is required at this stage in the development of multimodal studies as a field is the sort of empiricism of extensive text analysis. To do this inevitably require that analysts of multimodality learn to use (particularly contemporary digital) multimodal resources and techniques in order to appropriately deal with the natures of such media: to turn multimodal language back to itself (O`Halloran K.L, Smith B.A., *Multimodal Text Analysis*, 2-4, 7-11)

BIBLIOGRAPHY

DAVIES, M. & RAVELLI, L.J. (1992). Advances in Systematic linguistics: Recent Theory and Practice, London.

DOWNING, A. & LOCKE, P. (2006). English Grammar, a University course. New Yourk: Routledge.

Evaluative text analysis, first published in R. Steele and T.Threadgold, language Topics, Essays in Honour of Michael Halliday, 1987.

FROMKON, V. (2017) An Introduction to Language. USA: Cengage.

O`HALLORAN, K.L. & SMITH, B.A., "Multimodal Text Analysis", 2-4, 7-11.

POPPING, R., "Computer-Assisted Text Analysis", 1, 26-29.

PUAL, J. (1999). An Introduction to Discourse Analysis. Great Britain Clays Ltd.

SALKIR, R. (1995) Text and Discourse Analysis Great Britain: T.J. International Ltd, padstow, Cornwall.

TAUSCZIK, Y.R., PENNEBAKER, J. W. (2010). "The Psychological Meaning of Words: LIWC and Text Analysis Methods", Journal of Language and Social Psychology29(1) 24–54

YULE, G. (2010). The study of Language. New York: Cambridge press.

BIODATA

Emad Jasem Mohamed: Assistant Instructor of the English Department of the College of basic Education of Misan University, Maysan, Iraq. Imad.jasim@uomisan.edu.iq