

Kay L. O'Halloran, Sabine Tan, Peter Wignell, Rui Wang, Kevin Chai, and Rebecca Lange

Multimodality: A New Discipline

Abstract: Multimodality, involving the study of meaning arising from the integration of language with images and other resources in multimodal texts, interactions and events, addresses the fundamental need to understand human communication in the current age of digital technology. However, multimodality is not considered to be a discipline per se at present. By drawing parallels between mathematics and linguistics, it is proposed that if multimodality is to become a discipline, then abstract context-based frameworks for modeling multimodal semiotic resources and methodologies for investigating patterns of human communication are required. An example of how this could be achieved is provided. From here, multimodality has the potential to provide the foundations for a range of multimodal sciences, in much the same way that mathematics and linguistics underpin the mathematical and language sciences respectively. In doing so, it may become possible to track the changes in human communication arising from digital technology and the resultant impact on thought and reality.

Keywords: multimodality, discipline, mathematics, mapping, computational tools, context

1 Introduction

A discipline is generally understood to be an organized and systematic body of knowledge that is typically studied at university level. For example, the Oxford English Dictionary defines 'discipline' as "a branch of knowledge, typically one studied in higher education".¹ Much of multimodality as it is researched today evolved from branches of linguistics and social semiotics in the 1990s (e.g., Tan et al. 2019) as a means for studying the ways in which human beings use a whole range of different semiotic resources (including language, sound, gesture, images, and so forth) for meaning making and communication.

Today, multimodality can be viewed beneficially in the same light as mathematics and linguistics; namely, as a field that addresses a fundamental problem in

¹ <https://en.oxforddictionaries.com/definition/discipline>, last accessed: March 18, 2019.

contemporary society, in this case the need to understand human communication, particularly in the current age of digital technology (i.e., Internet, social media, and mobile devices). At the present time, multimodality is studied in various courses and has increasingly become the focus of postgraduate research. Despite these advances, however, multimodality is not considered to be a discipline per se at this stage. If multimodality is to become a discipline, it is proposed that generalizable, abstract context-based frameworks for modeling multimodal semiotic resources and analytical methods for investigating patterns of human communication over space and time are required.

In what follows, the ways in which multimodality may become a discipline in the future are explored by drawing parallels with mathematics and linguistics. These two fields developed in order to address key issues in the world—i.e., the modeling of the material world and the human world of language—and in doing so, developed abstract and generalized knowledge that could be applied in different real-life contexts. Firstly, mathematics flourished during the Renaissance when mathematical innovations in the form of abstract structures were linked to scientific discoveries. From there, mathematics became the science of number, quantity, and space for modeling and predicting the material world, giving rise to a range of mathematical sciences.

Secondly, modern linguistics turned to the study of grammatical systems in the 20th century to address key questions about human language in terms of language change, language structure, and language use. These developments gave rise to different language sciences informed by various branches of linguistics. From there, the two fields developed as disciplines that underpin a range of mathematical and language sciences respectively, as illustrated below. In the following sections, we discuss mathematics and linguistics respectively, before turning to multimodality.

2 Mathematics: Mapping the Physical World

Mathematics developed relatively slowly until mathematical innovations were linked to scientific discoveries in the Renaissance (e.g., Eves 1990). Galileo (1623 [1957]), as a pioneer of the scientific method, understood the significance of mathematics for advancing science:

“The universe cannot be read until we have learned the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which means it is humanly impossible to comprehend a single word. Without these, one is wandering about in a dark labyrinth.” (Galileo 1623 [1957], *Opere Il Saggiatore*)

Today, mathematics is described as “the abstract science of number, quantity, and space, either as abstract concepts (pure mathematics), or as applied to other disciplines such as physics and engineering (applied mathematics)”.² From this perspective, pure mathematics is viewed as mathematics for its own sake without any pre-determined applications, although applications are often found later on. On the other hand, applied mathematics is designed to solve specific problems, sometimes leading to new fields of mathematics (e.g., statistics and game theory). The simple division of mathematics into two categories is seen to create barriers, however, whereas in reality there are many commonalities across the mathematical sciences (National Science Council 2013).

Whichever way mathematics is considered, it is a human construction developed for certain purposes. As Kline (1980, 312) explains: “What then is mathematics if it is not a unique, rigorous, logical structure? It is a series of great intuitions carefully sifted, and organized by the logic men [sic] are willing and able to apply at any time”. In other words, mathematics is “a human construction with all that implies” (Little 1981, 159). Although mathematics is not an empirical science, many of the ideas originate in empirical results from which further concepts and areas are developed. As Neumann explains:

“I think that it is a relatively good approximation to truth—which is much too complicated to allow anything but approximations—that mathematical ideas originate in empirics, although the genealogy is sometimes long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better compared to a creative one, governed by almost entirely aesthetical motivations, than to anything else and, in particular, to an empirical science.” (Neumann 1956, 2063)

The National Science Council (2013) defines the mathematical sciences in broad terms: namely, those areas which “aim to understand the world by performing formal symbolic reasoning and computation on abstract structures” (National Science Council 2013, 62).

The various areas of the mathematical sciences are displayed in Figure 1. This includes the traditional areas (e.g., engineering, economics, computer science, geoscience, astronomy, physics, chemistry, and biology) and areas that are concerned with building mathematical models and exploring them computationally through the analysis of datasets (e.g., medicine, social networks, information processing, communications, defense, manufacturing, marketing, and finance). All these fields are mathematical in nature, regardless if they are part of computer science or part of the discipline for which the modeling or analysis are performed.

² <https://en.oxforddictionaries.com/definition/mathematics>, last accessed: March 18, 2019

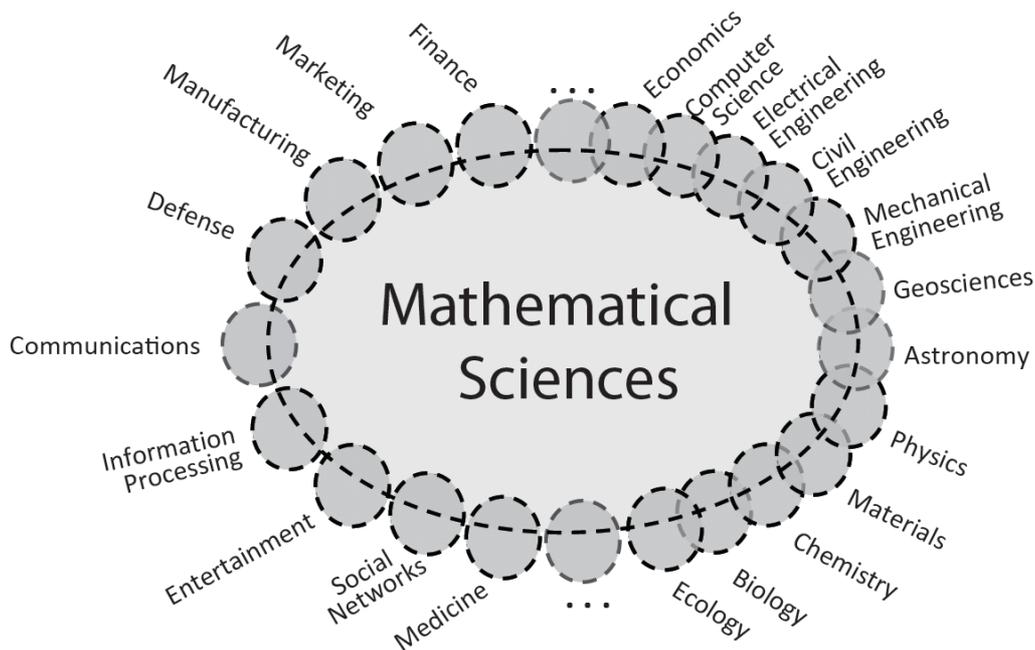


Fig. 1: The Mathematical Sciences (National Science Council, 2013, 63)

The various activities in the mathematical sciences aim to:

- discover relationships between the abstract structures;
- capture features of the world in abstract structures through modeling and formal reasoning or by using abstract structures as a framework for computation to make predictions about the world;
- use abstract reasoning, models, and structures to make inferences about the world through data science.

As the National Science Council explains, these activities are “linked to the quest to find ways to turn empirical observations into a means to classify, order, and understand reality—the basic promise of science” (National Science Council 2013, 62).

In a similar fashion, multimodality needs to be seen as originating in empirical results from which abstract concepts and ideas are formed in order to classify, order and understand human (rather than physical) reality (Bateman 2014a, O’Halloran et al. 2016, Tan et al. 2018). In this regard, we can view multimodality as a science with the potential to be applied to other areas, henceforth referred to as *the*

multimodal sciences. Before exploring these propositions further, we first consider linguistics as a discipline which shares similarities with mathematics and multimodality.

3 Linguistics: Mapping Human Language

Language has been an object of enquiry since antiquity; for example, logic, rhetoric, and grammar were studied in ancient Greece. Modern linguistics is the scientific study of language (Halliday 2003)³ which aims to answer key questions about human language in terms of language change, language structure, and language use. Linguistics, like mathematics, can be grouped into two main areas: pure (or theoretical, or general) linguistics, and applied linguistics, with various subfields in each category. However, many branches of linguistics do not fit easily into either category, given that they are concerned with developing theory in order to understand how language is used (e.g., systemic functional linguistics, psycholinguistics, sociolinguistics, and computational linguistics).

Following Halliday (1978), linguistics is primarily concerned with ‘language as system’ in terms of substance (phonic or graphic) and form (vocabulary, grammar and semantics), as shown in Figure 2 (see central triangle). In addition, linguistics is concerned with the study of ‘language as behavior’ (e.g., socialization and sociolinguistics), ‘language as knowledge’ (psycholinguistics), and ‘language as art’ (e.g., literary studies) (see Figure 2). The different areas of linguistics are related to other disciplines: for example, sociology, psychology, literature, and physics and physiology. Beyond this, linguistics is involved in other areas such as archeology, philosophy, logic and mathematics, communications engineering, culture, social anthropology, for example.

Linguistics mirrors mathematics in that core areas provide the basis for other areas of study, in this case, the language sciences. In order for this to occur, it was necessary to develop abstract structures to explain how language works as a system. For example, Halliday (1973) describes the grammatical systems through which language fulfills certain functions (see Figure 3), which he later developed into a comprehensive lexicogrammar of English (e.g., Halliday & Matthiessen 2013). In Halliday’s model, the grammatical systems are organized into ranks according to three metafunctions: (a) *ideational meaning* consisting of experiential meanings to capture happenings in the world, and logical meaning to capture the logical relations between those happenings; (b) *interpersonal meaning* to map social

³ <https://en.wikipedia.org/wiki/Linguistics>, last accessed: March 18, 2019

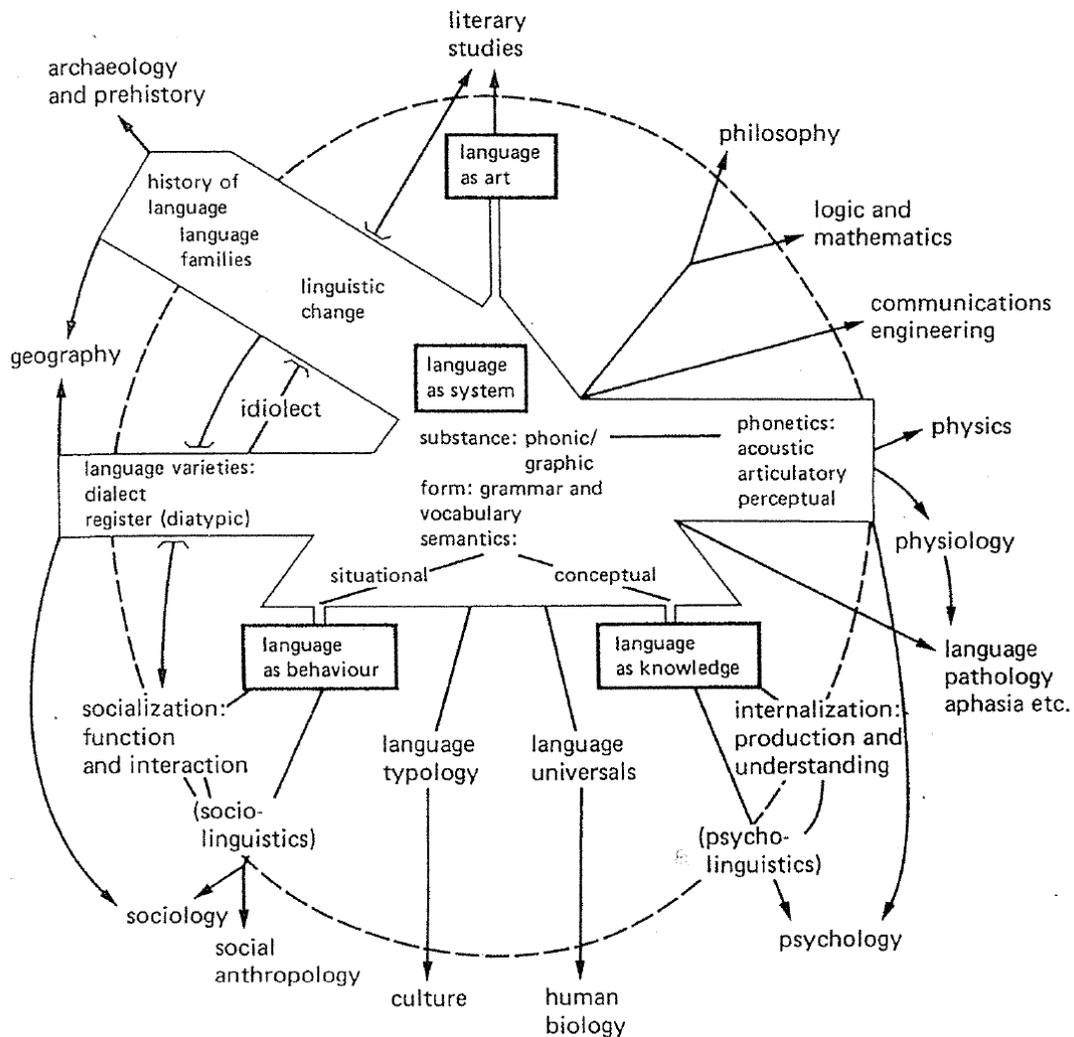


Fig. 2: Domains of language studies and their relations to other fields (Halliday 1978, 11)

interactions and relations; and (c) *textual meaning* to organize the message. The grammatical systems are used to map choices in linguistic texts and to examine relations between system choices according to roles which language is playing in different contexts.

The meaning of human language is dependent on context, however, so that the mapping between language choices and meaning is not straightforward. For this reason, it has not been possible to implement computational models for the full range of linguistic systems (e.g., see discussion in Bateman & O'Donnell 2015). Indeed, most natural language processing algorithms focus on lexical items, rather than grammatical systems which underlie the functional organization of language. However, the potential for modeling context has been greatly enhanced in the digital age, due to the large datasets of text, images, and videos which are now

rank		metafunction		ideational			experiential		intrapersonal		textual	
				logical			TRANSITIVITY (process type)	MOOD MODALITY POLARITY	THEME CULMINATION VOICE	(cohesive)		
clause		comp lexes (clause -			TRANSITIVITY (process type)	MOOD MODALITY POLARITY	THEME CULMINATION VOICE	COHESIVE RELATIONS				
phrase	[prepositional]	phrase -			MINOR TRANSITIVITY (circumstance type)	MINOR MOOD (adjunct type)	CONJUNCTION					
group	[verbal]	group -	INTERDEPENDENCY (parataxis/hypotaxis)	TENSE	EVENT TYPE ASPECT (non-finite)	FINITENESS	VOICE DEICTICITY	REFERENCE ELLIPSIS/SUBSTITUTION				
	[nominal]		&		THING TYPE CLASSIFICATION	PERSON ATTITUDE	DETERMINATION	CONJUNCTION				
	[adverbial]		LOGICAL-SEMANTIC RELATION (expansion/projection)	MODIFICATION	EPITHESES QUALIFICATION	COMMENT (adjunct type)	CONJUNCTION					
word		word)		MODIFICATION	QUALITY (circumstance type)	(CONNOTATION)						
				DERIVATION	(DENOTATION)							
information unit		info. unit complex	complexes	ACCENTUATION		KEY	INFORMATION					
			simplexes									

A "function / rank matrix" for the grammar of English, where (i) rows show rank and primary class,

(ii) columns show metafunction, and (iii) capitals show system(s) in each cell

Fig. 3: Halliday's function/rank matrix for the grammar of English (Halliday 1973, 141)

available with metadata (e.g., URLs, date of postings, source materials, and references). In addition, semantic categorizations of (nearly) every domain of human activity are now available through socially evolved knowledge classification systems such as Wikipedia. Therefore, it is possible to analyze human communication in relation to context in new ways that have not existed before, given the metadata for large datasets of multimodal texts which is available today.

For this reason, we propose that multimodality is poised ready to follow a similar trajectory to mathematics and linguistics in terms of developing generalizable, abstract structures that can be applied to different contexts of communication in the real world. If this does occur, then multimodality will become a scientific discipline, giving rise to the multimodal sciences for modeling and mapping the human universe. The multimodal data is available now, but we do not have the necessary abstract models and methodologies yet. In what follows, a possible path forward in this direction is discussed.

4 Multimodality: Mapping the Human Universe

As mentioned above, much of modern multimodality originated in linguistics, particularly in social semiotics and systemic functional linguistics (Bateman et al. 2017, Jewitt 2014, Jewitt et al. 2016, Tan et al. 2019). Multimodality is concerned with the entire range of semiotic resources which humans use for meaning making, including language, image, symbolism, gaze, gesture, space, architecture, and so forth. In particular, multimodality is concerned with the integration of language with other systems of meaning and mapping the interaction of semiotic choices in texts, interactions and events in different contexts. Tan et al. (2019) provide a comprehensive account of recent theoretical, methodological, and analytical trends in multimodality, and this review is not repeated here. Rather, the focus of this discussion is how multimodality may become a science which provides the foundations for other fields of study, with the leading question: “what are the requirements for multimodality in terms of mapping the human world?”.

Mathematics succeeded by providing semiotic tools for formulating abstract structures, which could be used for modeling and formal reasoning, and as frameworks for computation to make predictions about the physical world. The interrelations between concepts, systems, and processes are made explicit, reasoning and computation are made as efficient as possible, and the limits of the findings are characterized. Moreover, the abstract structures hold, regardless of the context in which they are applied and used in the physical world. The definition of ‘abstract structure’ makes this clear:

An abstract structure may be represented (perhaps with some degree of approximation) by one or more physical objects—this is called an implementation or instantiation of the abstract structure. *But the abstract structure itself is defined in a way that is not dependent on the properties of any particular implementation.* [emphasis added]⁴

Similarly, (Halliday 2008, 7) views systemic functional linguistics as an “applicable science” with a “comprehensive and theoretically powerful model of language” designed to address problems associated with language use. The systemic functional model of language incorporates context (i.e., the context of situation and the context of culture derived from Malinowski (1923)) but it has not been possible to formalize contextual parameters to the same extent as the lexicogrammar as yet. As a result, instances of language use could not be fully accounted for, given that the meanings of linguistic choices arise from their context of use.

Nonetheless, Halliday’s systemic functional model of language is a comprehensive description of how language is organized to create meaning (as a system of meanings), and how these systems are activated to fulfill certain functions in relation to context. Significantly, the basic principles of language as a social semiotic system can also be applied to images, videos, and other resources, resulting in frameworks with common theoretical concepts of metafunctions, systems, and ranks (e.g., see Figure 4), despite the different resources which are involved. This provides a common foundation upon which to model and analyze different semiotic resources in terms of their underlying organization in the form of metafunctionally based systems, organized according to different ranks, as displayed in Figure 4.

Following Halliday (2008), multimodality is conceptualized an “applicable science” that is designed to address problems associated with the use of language, images, and other semiotic resources. Furthermore, it is proposed that *abstract context-based models of semiotic resources* and *semiotic interactions* are required in order to map patterns of meaning in human communication, and to trace those patterns over space and time. These abstract models are designed to

- discover relations between semiotic resources;
- map patterns of semiotic choices in texts, interactions, and events;
- provide an overarching framework for computational models for mapping patterns and making predictions about the human world. This includes making inferences through reasoning, and models and structures using data science.

⁴ https://en.wikipedia.org/wiki/Abstract_structure, last accessed: March 18, 2019.

Genre			
Register (Field, Tenor and Mode)			
Language			Text/Image
Experiential	Rank	System	Description
	Clause	Processes; Participant Roles; Circumstance	Happenings, actions and relations
	Clause	Logico-Semantic Relations	Relations between happenings, actions and relations
	Complex Discourse	Appraisal	Evaluation in terms of attitude, emotion and judgment
	Clause	Speech Function	Exchange of information (e.g. statements and questions) and goods & services (e.g. commands and offers)
Textual	Clause	Information Focus	Organisation of information, with points of departure for what follows
	Discourse		
Images			
Experiential	Rank	System	Description
	Work	Narrative Theme; Representation; Setting	Nature of the scene
	Episode	Processes; Participant Roles; and Circumstance	Visual happenings, actions and relations
	Figure	Posture; Dress	Characteristics of the participants
	Work	Logical Relations	Relations between process and participant configurations (e.g. temporal, spatial, causal)
	Episode	Logical Relations	Relations between participants
	Figure	Logical Connections	Relations between parts of figures and objects
	Work	Angle; Camera Distance; Lighting	Visual effects
	Episode	Proportion in Relation to the Whole Image; Focus; Perspective	Happenings, actions and relations with respect to the whole image
	Figure	Gaze-Visual Address	Direction of participant's gaze as internal to image or external to viewer
Textual	Work	Compositional Vectors; Framing	The organisation of the parts as a whole, with the visual marking (e.g. framing) of certain parts
	Episode	Relative Placement of Episode; Framing	Position of the happenings, actions and relations in relation to the whole image, and the visual marking of certain aspects
	Figure	Relative Placement of the Figure within the Episode; Arrangement; Framing	Position of figures in relation to happenings, actions or relations, and the visual marking of certain aspects of those figures

Fig. 4: Language and Images (e.g., O'Halloran et al. 2016)

The scientific view of multimodality mirrors that of mathematics (for example, as formulated by the U.S. National Science Council (2013, 62)) and linguistics (for example, as developed by Michael Halliday) with the goal of mapping the human world that includes and extends beyond language. Rephrasing the National Science Council (2013, 62), “this is linked to the quest to find ways to turn empirical observations into a means to classify, order, and understand *human* reality—the basic promise of *multimodality*”. However, there are two major problems: first, mathematics deals with abstract structures which are independent of applications in the real world (unlike language and multimodality which are context-dependent), and secondly, semiotic formulations beyond language are required in this model. These issues are discussed below.

5 Multimodality: Context and Semiotic Resources Beyond Language

Mathematics is the study of abstract structures, which are defined by laws, properties, and relationships which hold, regardless of the context. That is, the abstract structures can be represented in the physical world but the abstract structures themselves are independent of the properties of any particular instantiation. Computer language, for example, is considered to be an abstract structure because it can be implemented with the same result in any context, but natural language is not generally perceived to be an abstract structure because it can be used with different results according to the context of use. For example, “I like it” can mean different things (i.e., ‘I really do like it’, or ‘I really don’t like it’) according to how it is said and/or written and the context of the use. This same argument applies to multimodal texts, but the problem is exacerbated because the meaning arises from combinations of interacting semiotic choices which are interpreted in relation to the context.

Therefore, in order for language, images, and other resources and semiotic interactions to be modeled as abstract structures, these abstract structures need to incorporate context to account for the variations in meaning which occur in the instantiations of multimodal choices. But how can this be done? The problem is foregrounded in context-enhanced information fusion (Snidaro et al. 2016) where context is taken into account at different levels of abstraction; for example, the low-level data, feature extraction, patterns between features, and decisions and relationships and high level descriptions, as displayed in Figure 5. Furthermore, horizontal and vertical heterogeneity are incorporated in the model, as displayed in Figure 5. In what follows, we discuss possible approaches to modeling and

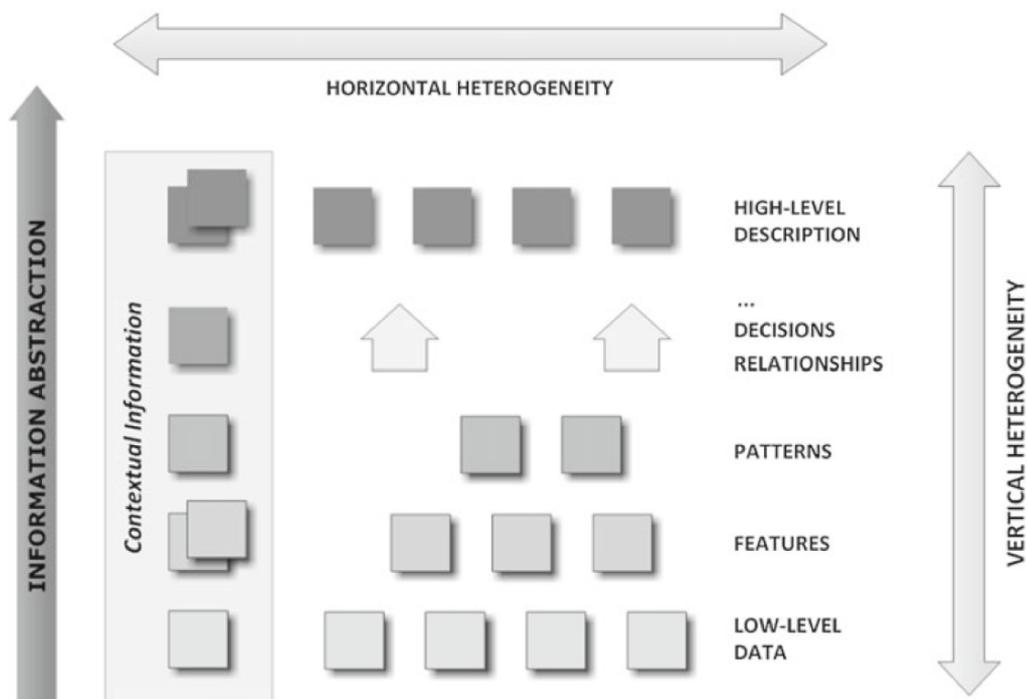


Fig. 5: Context-Enhanced Information Fusion Snidaro et al. (2016, 434)

analyzing low level data, features, and patterns, while recognizing that decisions regarding relationships and high-level descriptions involve further contextual parameters (e.g., decisions involving the emerging patterns and high-level descriptions, such as prediction).

One possible way forward is to use computational models for language, images, videos, and other resources because these models are already formulated in terms of abstract structures. However, computational models were typically developed for one resource: for example, natural language processing, image processing, and video processing. Moreover, these computational approaches identify low-level data features and use machine learning (e.g., neural networks) to identify lexical items in written texts and objects and events in images and videos.

Recent developments in information fusion aim to combine different modalities (e.g., Arevalo et al. 2017, Kiela et al. n.d.). However, three major challenges exist: namely, “feature learning and extraction, modeling of relationships between data modalities and scalability to large multimodal collections” (Arevalo 2018, 1). Given this situation, multimodality offers an exciting opportunity to contribute to this field, as evidenced by work which is underway (see overview in Bateman et al. 2019). In what follows, we describe how various computational tools can

be integrated into a multimodal framework for big data analytics of multimodal communications.

6 Integrating Computational Tools with Multimodal Theory and Context

In order to formulate multimodal systems and choices as generalizable abstract structures which can be applied to any instance of use, it is proposed that state-of-the-art automated computational techniques (e.g., text, image, and video processing) are embedded in a multimodal framework which incorporates contextual parameters provided by various forms of metadata. Indeed, research efforts along these lines are already underway.

For example, mixed methods approaches involving the integration of multimodal analysis, data mining, and information visualization for modeling patterns of multimodal communications in large-scale data have been proposed (e.g., O’Halloran et al. 2016, Tan et al. 2018). The computational techniques provide the necessary abstract structures and foundations for scientifically investigating multimodal discourses across media platforms and contexts, building a basis for the future development of multimodality as a discipline.

As a further step forward with this initiative, an approach which incorporates automated computational techniques (e.g., text, image and video processing algorithms) within a multimodal framework and uses machine learning techniques for the analysis of big multimodal datasets is proposed. In this approach, the basic methodology involves representing the multiple dimensions of multimodal texts (e.g., language, image, video, and context) using vectors which record the presence or absence of each feature. Machine learning algorithms are then used to classify multimodal texts according to ideational formulations which are realized and the multimodal strategies which are used. As displayed in Figure 6, the approach has four steps: 1. Feature Extraction; 2. Multimodal Feature Enhancement; 3. Feature Representation; and 4. Classification. These steps are explained in turn below.

1. Feature Extraction: Natural language understanding (NLU) algorithms (for example, IBM Watson) are applied to the language components of the multimodal texts. For example, the NLU models in IBM Watson are *categories, concepts, emotion, entities, keywords, metadata, relations, semantic roles, sentiment, metadata* and *relations*⁵. General descriptions of the language models are provided by the

⁵ <https://natural-language-understanding-demo.ng.bluemix.net/>, last accessed: March 18, 2019

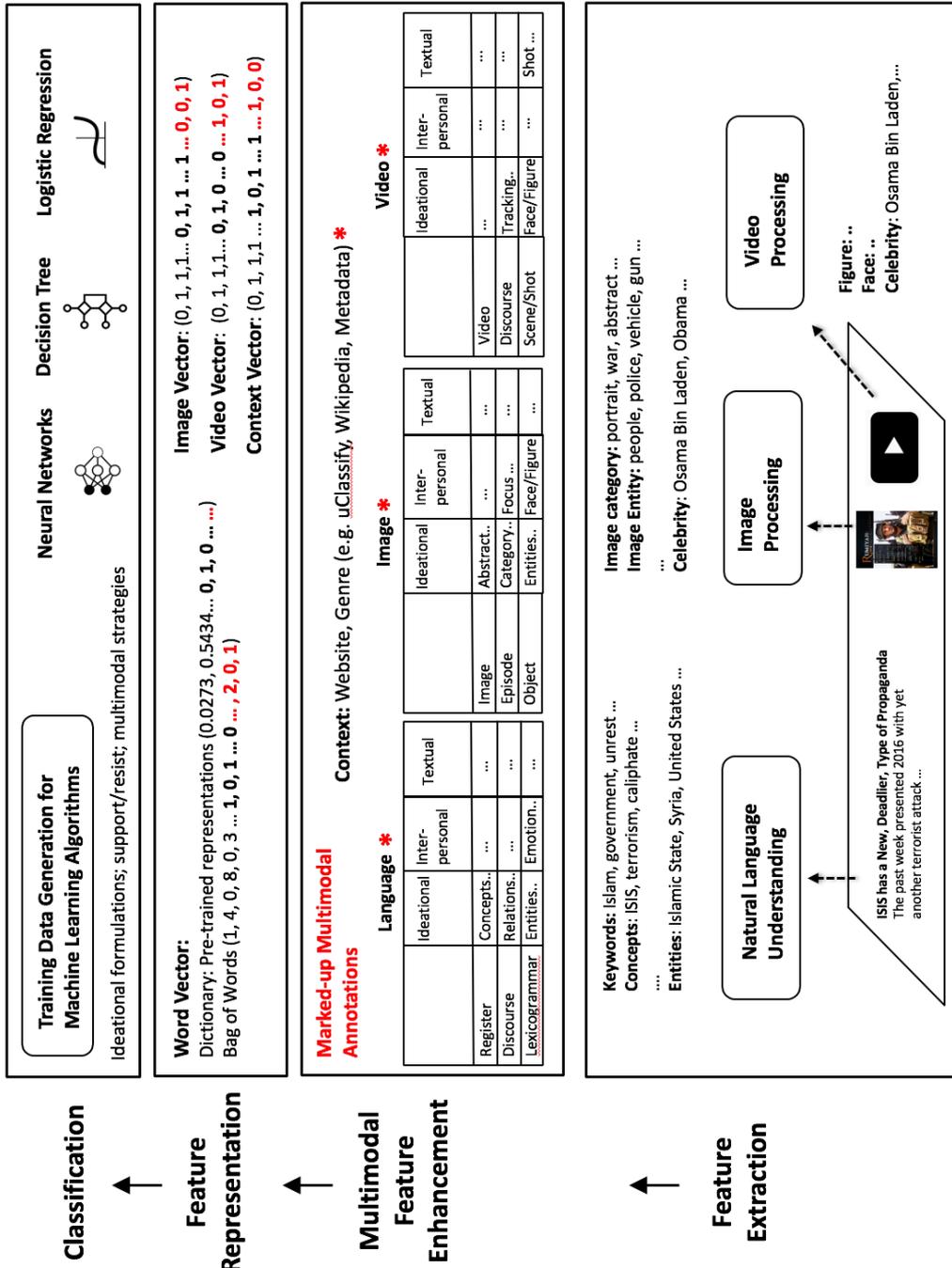


Fig. 6: Multimodality as Abstract Structures

developers of the NLU tools, but as with most commercial NLU models, the criteria, rationale, and algorithms are not provided so these models are black boxes. Nonetheless, the usefulness of the language models when integrated in a multimodal analysis framework, supplemented by text tagging, have been demonstrated (Wignell et al. 2018).

Similarly, visual processing models (for example, DenseCap⁶, Clarifai⁷, Google Cloud Vision⁸, and IBM Watson Visual Recognition⁹) are applied to the images in the multimodal texts to extract semantic information.

For example, the image processing models include *object labelling*, *object bounding*, *face detection*, *face bounding*, *face analysis*, *logo detection*, *logo bounding*, *celebrity detection*, *celebrity bounding*, *apparel labelling* and *web detection*. These image models are also black boxes which nonetheless have proved useful for extracting information from the images (Cao & O'Halloran 2015, O'Halloran et al. 2014, Podlasov & O'Halloran 2014). In addition, video processing models (e.g., Amazon Rekognition¹⁰) are applied to extract information from the videos in the multimodal texts. The models include *object*, *scene*, and *activity detection*, *facial recognition*, *facial analysis*, *pathing*, *celebrity recognition* and *text-in-image recognition* for identifying participants, objects, events, and text in the videos.

2. Multimodal Feature Enhancement: The various computational models for language, images, and videos are integrated within a multimodal analysis framework, so that each model is categorized according to semiotic resource, meta-function (experiential, logical, interpersonal, and textual), and rank, as displayed in Figure 6. This means that the results from the computational models are marked up according to the multimodal theoretical framework, which also indicates the gaps where there is missing information. For example, NLU models are largely concerned with experiential meaning at the rank of word group, and neglect the textual organization, where certain elements have a greater semantic input due to the functions of those element (e.g., headline, caption, lead paragraph) (Wignell et al. 2018).

Furthermore, the multimodal feature enhancement takes context into account by incorporating metadata (e.g., URLs, date of postings, source materials, and references) so that the multimodal texts are annotated according to location, time, text type, and other attributes. For example, the text and URLs of websites are

⁶ <https://cs.stanford.edu/people/karpathy/densecap/>, last accessed: March 18, 2019

⁷ <https://clarifai.com/>, last accessed: March 18, 2019

⁸ <https://cloud.google.com/vision/>, last accessed: March 18, 2019

⁹ <https://www.ibm.com/watson/services/visual-recognition/>, last accessed: March 18, 2019

¹⁰ <https://aws.amazon.com/rekognition/>, last accessed: March 18, 2019

analyzed using various algorithms (e.g., uClassify¹¹) which classifies the website into different types of news stories and topics. In this way, multimodal theory and the context are incorporated in the descriptions of the multimodal texts.

3. Feature Representation: The various features of the multimodal analysis and the context are represented by a series of vectors. Dummy variables are used for representing images, videos, and context, i.e. using “0” and “1” to indicate the absence and presence of a feature respectively. Texts are represented using the bag-of-words representation, where a text is a dictionary vector and values corresponds to frequencies of the words appearing in the text. In addition to bag-of-words, texts can be also represented as word embeddings for training neural networks.

Word embeddings provide a series of dense, real-number vectors that are pre-trained over large amount of texts, e.g., a Wikipedia snapshot, which enables embedding vectors to encode semantics of words (Mikolov et al. 2013, Pennington et al. 2014). By coupling representations of images, videos, context, and texts, the multiple dimensions of the multimodal analysis (semiotic resource, metafunction, system, rank, and context) are incorporated into the model, resulting in a multidimensional description of the features of the multimodal texts, consisting of thousands (or more) dimensions. In this way, the complexity of the multimodal analysis is accounted for in the approach.

4. Classification: The multimodal texts are classified using machine learning algorithms which have been trained using previously classified data. Examples of machine learning algorithms include neural networks, decision trees, logistic regression, and other statistical methods. For example, K-modes clustering (Huang 1997, 1998) and an interactive visualization application are being used to analyze the reuse of images from online terrorist propaganda across different media platforms. The multimodal texts are clustered according to similarities and differences derived from the vectors with the list of features for each multimodal text (see O'Halloran et al. 2016). The proposed techniques and methodologies for integrating multimodal theory with computational models for text, image, videos, and contextual information result in large-scale mapping of the semantic space of multimodal texts, together with classifications of the ideational formations which are created and the multimodal strategies which are used.

In addition, it is necessary to display the results using some form of interactive visualization in order to explore the results (O'Halloran et al. 2016, Tan et al. 2018) The methodology presented here is currently being tested with real-life data and an interactive visualization explores the usefulness of the approach for analyzing large datasets of multimodal texts. That is, the approach is being empirically tested. This

¹¹ <https://www.uclassify.com/browse>

is the basic proposition advocated in the current discussion: i.e. to test concepts, systems, and processes in a rigorous fashion so that the relations are made explicit, reasoning and computation are made as efficient as possible, and the limits of the findings are specified. As such, this discussion presents possible steps towards the scientific study of multimodality, with a view to paving the way for the multimodal sciences.

7 Conclusions

Multimodality holds great promise for addressing serious problems in the world today where truth itself is at stake, given the current era where private corporations are employed to spread false information to influence the outcome of political processes. If multimodality continues to develop by building abstraction upon abstraction without an empirical basis, the result will be “abstract inbreeding”, leading to the possible degeneration of the field (see discussion of mathematics in Neumann 1956, 2063). Indeed, Bateman and colleagues (Bateman 2014*a,b*, 2016, Bateman et al. 2004, 2019) have also called for an empirical basis to multimodal research in order to provide firm foundations for the future development of the field.

Looking back, mathematics and linguistics developed in order to address specific problems at the time. From here, each area developed into disciplines which provided the basis for mathematical and language sciences respectively. In much the same way, multimodality has the potential to address key issues of human communications, leading to a range of multimodal sciences. Recent studies in the United States reveal that there is an increasing number of students completing degrees in linguistics, particularly at undergraduate level, although the rate has slowed down in recent years (The Linguistic Society of America 2017). The increased interest may be related to the changes in the communication landscape resulting from digital technology. Whatever the reason, this trend offers a promising scenario for multimodality which moves beyond the study of language in isolation to the study of language as it combines with other semiotic resources in human communication.

Galileo’s view of the critical role of mathematics for understanding and predicting the physical world can be extended to multimodality in terms of providing tools for understanding the human world. Paraphrasing Galileo makes this connection explicit:

The human universe cannot be read until we have learned the language and become familiar with the characters in which it is written. It is written in *multimodal language*, and the *signs* are *multidimensional in nature*, without which means it is humanly impossible to comprehend a single *dimension*. Without these, one is wandering about in a dark labyrinth. Based on Galileo (1623 [1957], *Opere Il Saggiatore*).

If multimodality becomes a discipline which provides the foundations for the multimodal sciences, it may become possible to understand the changes in human communication arising from digital technology and the resultant impact on thought and reality.

Bibliography

- Arevalo, J. (2018), *Multimodal Representation Learning with Neural Networks.*, PhD thesis, Engineering School, Systems and Industrial Engineering Department, National University of Colombia.
- Arevalo, J., Solorio, T., y Gomez, M. M. & Gonzalez, F. (2017), Gated multimodal units for information fusion, *in* 'Proceedings of the Workshop of the 5th International Conference on Learning Representations (ICLR)', Workshop of the 5th International Conference on Learning Representations (ICLR).
<https://arxiv.org/pdf/1702.01992.pdf>
- Bateman, J. A. (2014a), Looking for what counts in film analysis: a programme of empirical research, *in* D. Machin, ed., 'Visual Communication', Mouton de Gruyter, Berlin, pp. 301–330.
- Bateman, J. A. (2014b), Using Multimodal Corpora for Empirical Research, *in* C. Jewitt, ed., 'The Routledge Handbook of multimodal analysis', 2 edn, Routledge, London, pp. 238–252.
- Bateman, J. A. (2016), Methodological and theoretical issues for the empirical investigation of multimodality, *in* N.-M. Klug & H. Stöckl, eds, 'Sprache im multimodalen Kontext / Language and Multimodality', number 7 *in* 'Handbooks of Linguistics and Communication Science (HSK)', de Gruyter Mouton, Berlin, pp. 36–74.
- Bateman, J. A., Delin, J. L. & Henschel, R. (2004), Multimodality and Empiricism: Preparing for a Corpus-based Approach to the Study of Multimodal Meaning-Making, *in* E. Ventola, C. Charles & M. Kaltenbacher, eds, 'Perspectives on Multimodality', John Benjamins, Amsterdam, pp. 65–87.
- Bateman, J. A., McDonald, D., Hiippala, T., Couto-Vale, D. & Costechi, E. (2019), Systemic-functional linguistics and computation: New directions, new challenges, *in* G. Thompson, W. L. Bowcher, L. Fontaine, J. Y. Liang & D. Schönthal, eds, 'The Cambridge Handbook of Systemic Functional Linguistics', Cambridge University Press, Cambridge UK.
- Bateman, J. A. & O'Donnell, M. (2015), Computational Linguistics: the Halliday Connection, *in* J. J. Webster, ed., 'The Bloomsbury Companion to M.A.K. Halliday', Bloomsbury, London and New York, pp. 453–466.
- Bateman, J. A., Wildfeuer, J. & Hiippala, T. (2017), *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*, Mouton de Gruyter, Berlin.

- Cao, Y. & O'Halloran, K. L. (2015), 'Learning human photo shooting patterns from large-scale community photo collections', *Multimedia Tools and Applications* 74(24), 11499–11516.
- Eves, H. (1990), *An Introduction to the History of Mathematics*, Saunders College, New York.
- Galileo, G. (1623 [1957]), *Opere il saggiaatore*, in 'Discoveries and Opinions of Galileo', Doubleday and Company, pp. 237–238.
- Halliday, M. A. K. (1973), *Explorations in the Functions of Language*, Edward Arnold, London.
- Halliday, M. A. K. (1978), *Language as social semiotic*, Edward Arnold, London.
- Halliday, M. A. K. (2003), *Collected Works of M. A. K. Halliday. Volume 3. On Language and Linguistics*, Continuum, London and New York.
- Halliday, M. A. K. (2008), Working with meaning: Towards an applicable linguistics, in J. J. Webster, ed., 'Meaning in Context: Strategies for Implementing Intelligent Applications of Language Studies', Continuum, London, pp. pp. 7–23.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2013), *Halliday's Introduction to Functional Grammar*, 4 edn, Routledge, London and New York.
- Huang, Z. (1997), Clustering large data sets with mixed numeric and categorical values, in Hiroshi Motoda & Hongjun Lu, eds., in 'Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference, Singapore', World Scientific Publishing Co Pte Ltd, Singapore, pp. 21–34.
- Huang, Z. (1998), 'Extensions to the k-modes algorithm for clustering large data sets with categorical values', *Data Mining and Knowledge Discovery* 2(3), 283–304.
- Jewitt, C., Bezemer, J. & O'Halloran, K. (2016), *Introducing multimodality*, Routledge, London.
- Jewitt, C., ed. (2014), *The Routledge Handbook of Multimodal Analysis*, Routledge, London and New York. 2nd edition.
- Kiela, D., Grave, E., Joulinand, A. & Mikolov, T. (n.d.), Efficient large-scale multi-modal classification, in 'Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)', AAAI Press, Palo Alto, California USA, pp. 5198–5204.
- Kline, M. (1980), *Mathematics: The Loss of Certainty*, Oxford University Press, New York.
- Little, J. (1981), 'Review: Mathematics: The loss of certainty', *New Scientist*, pp. 159–159.
- Malinowski, B. (1923), *The problem of meaning in primitive languages*, Harcourt, Brace, and Co., Inc., pp. 451–510. Supplement I to C.K. Ogden and I.A. Richards *The Meaning of Meaning*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, G. (2013), Distributed representations of words and phrases and their compositionality, in C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K. Q. Weinberger, eds, 'Advances in Neural Information Processing Systems 26 (NIPS 2013)', NIPS Foundation, Lake Tahoe, Nevada, USA., pp. 3111–3119.
- National Science Council (2013), *The Mathematical Sciences in 2025*, The National Academies Press, Washington D.C.
- Neumann, J. (1956), The mathematician, in J. R. Newman, ed., 'The World of Mathematics', Vol. 4, Dover Publications Inc, New York, pp. 2053–2063.
- O'Halloran, K. L., Chua, A. & Podlasov, A. (2014), The role of images in social media analytics: A multimodal digital humanities approach, in D. Machin, ed., 'Visual Communication', Mouton de Gruyter, Berlin, pp. 565–588.
- O'Halloran, K. L., Tan, S., Pham, D.-S., Bateman, J. A. & Vande Moere, A. (2016), 'A Digital Mixed Methods Research Design: Integrating Multimodal Analysis with Data Mining and Information Visualization for Big Data Analytics', *Journal of Mixed Methods Research*.

- Pennington, J., Socher, R. & Manning, C. D. (2014), Glove: Global vectors for word representation, in 'Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)', Association for Computational Linguistics, Doha, Qatar, pp. 1532–1543.
- Podlasov, A. & O'Halloran, K. L. (2014), Japanese street fashion for young people: A multimodal digital humanities approach for identifying socio-cultural patterns and trends, in E. Djonov & S. Zhao, eds, 'Critical Multimodal Studies of Popular Culture', Routledge, London and New York, pp. 71–90.
- Snidaro, L., Garcia, J., Llinas, J. & Blasch, E. (2016), *Context-Enhanced Information Fusion: Boosting Real-World Performance with Domain Knowledge*, Springer International Publishing AG Switzerland, Cham Switzerland.
- Tan, S., O'Halloran, K. L. & Wignell, P. (2019), Multimodality, in A. D. Fina & A. Georgakopoulou, eds, 'Handbook of Discourse Studies', Cambridge University Press, Cambridge.
- Tan, S., O'Halloran, K. L., Wignell, P., Chai, K. & Lange, R. (2018), 'A multimodal mixed methods approach for examining recontextualisation patterns of violent extremist images in online media', *Discourse, Context and Media* 21, 18–35.
- The Linguistic Society of America (2017), The State of Linguistics in Higher Education: Annual Report, Technical report, The Linguistic Society of America.
https://www.linguisticsociety.org/sites/default/files/Annual_Report_2017_Final_2.pdf
- Wignell, P., Chai, K., Tan, S., O'Halloran, K. L. & Lange, R. (2018), 'Natural language understanding and multimodal discourse analysis for interpreting extremist communications and the re-use of these materials online', *Terrorism and Political Violence* .
<https://doi.org/10.1080/09546553.2018.1520703>