EXPLORING THE DATA TURN OF PHILOSOPHY OF LANGUAGE IN THE ERA OF BIG DATA

Shasha Xu

Ph. D. Associate Professor. School of Foreign Languages, Zhejiang University
of Finance & Economics, Hangzhou, 310018 – China.
https://orcid.org/0000-0003-0597-3517 | xushasha@zufe.edu.cn

Qian Yang

School of Foreign Languages, Zhejiang University of Finance & Economics, Hangzhou, 310018 – China.
https://orcid.org/0000-0001-5422-7022 | yang_qian@zufe.edu.cn

Abstract: Collection of data in our Information Technology age caused a revolution in knowledge. The unprecedented growth of data in the big data era has necessitated changes in the scale, nature, and status of data, leading researchers to adopt new paradigms and methodologies in philosophical research. In particular, the theoretical focus of philosophy of language shifted towards cognitive knowledge, with an emphasis on the proposition of data turn in cognitive cognition in the era of big data. The paper explores the potential scope for quantitative research on the data turn of philosophy of language by examining the need for transforming qualitative and quantitative research paradigms, by reconstructing the quantitative approach to philosophy of language, and by expanding human-data relations in the philosophy of big data. The paper concludes that there is the necessity for further research to examine the relationship between language, data and philosophy.

Keywords: Philosophy of language. Big data. Data turn. Quantitative analysis.

Abstract: La raccolta di dati nella nostra era dell'"Information Technology" ha generato una rivoluzione nella conoscenza. Nell'era dei "big data", la conseguente crescita senza precedenti dei dati, ha reso necessari cambiamenti nella scala, nella natura e nello stato dei dati, portando quindi i ricercatori ad adottare nuovi paradigmi e metodologie nella ricerca filosofica. In particolare, l'attenzione teorica della filosofia del linguaggio si è spostata verso la conoscenza cognitiva, con un'enfasi sulla proposizione particolare del "data turn" nella cognizione cognitiva nell'era dei "big data". Il paper esplora la potenziale portata della ricerca quantitativa del "data turn" nella filosofia del linguaggio, tramite l'analisi della necessità di trasformare i paradigmi della ricerca qualitativa e quantitativa, ricostruendo l'approccio quantitativo della filosofia del linguaggio ed ampliando l'analisi delle relazioni uomo-dati nella filosofia dei "big data". Il paper conclude affermando che sono necessarie ulteriori ricerche per esaminare in modo ancor più approfondito la relazione tra linguaggio, dati e filosofia.

Parole chiave: Filosofia del linguaggio. Big data. Data turn. Analisi quantitative.

Exploring the data turn of philosophy of language in the era of big data[1]

Shasha Xu [2]

Qian Yang [3]

Abstract: Collection of data in our Information Technology age caused a revolution in knowledge. The unprecedented growth of data in the big data era has necessitated changes in the scale, nature, and status of data, leading researchers to adopt new paradigms and methodologies in philosophical research. In particular, the theoretical focus of philosophy of language shifted towards cognitive knowledge, with an emphasis on the proposition of data turn in cognitive cognition in the era of big data. The paper explores the potential scope for quantitative research on the data turn of philosophy of language by examining the need for transforming qualitative and quantitative research paradigms, by reconstructing the quantitative approach to philosophy of language, and by expanding human-data relations in the philosophy of big data. The paper concludes that there is the necessity for further research to examine the relationship between language, data and philosophy.

Keywords: Philosophy of language. Big data. Data turn. Quantitative analysis.

Introduction

From the last decades of the 19th century, philosophers became more and more attentive to the role language plays in producing knowledge. While traditional topics continued to be debated by means of traditional methodologies, a mainstream focus on linguistic reflections has progressively emerged as peculiar feature of these times, to the extent that researchers qualify the way of making philosophy after 1950 as *linguistic turn* (Rorty, 1967; Losonsky, 2006). The characteristic feature of the linguistic turn consists in approaching normativity without relying on a foundational perspective. The key move was a massive use of mathematical logic as a tool to the purpose of rendering philosophical inquiry rigorous, clear, and conceptually explicit (Rorty, 1967). Nevertheless, logical analysis and reconstruction of metaphysical language proved to be challenging. Moreover, the logical analysis of natural language, conducted by language philosophers, leaves it

[2] Ph. D. Associate Professor. School of Foreign Languages, Zhejiang University of Finance & Economics, Hangzhou, 310018 – China. ORCID: https://orcid.org/0000-0003-0597-3517. E-mail: xushasha@zufe.edu.cn.

[3] School of Foreign Languages, Zhejiang University of Finance & Economics, Hangzhou, 310018 – China. ORCID: https://orcid.org/0000-0001-5422-7022. E-mail: yang_qian@zufe.edu.cn.

fragmented. The linguistic turn has failed to achieve its original academic vision, despite many notable achievements (Dummett, 2014).

In this context, the most relevant event to the future of philosophy of language (and philosophy in general) resides in that the world has now entered a new era of information revolution driven by the Internet, big data and artificial intelligence. Technical developments in natural and social sciences, including artificial intelligence, human-computer interaction and the Internet, all rely heavily on big data. As a result, data has become a central focus, and dataization has emerged as a fundamental component of the information revolution. In response to the change, philosophical research has undergone a series of changes, including research objects, paradigms and methodologies (Ji; Li; Qiu *et al.* 2012, p. 1; Furner, 2017, p. 55). Understanding philosophical questions requires consideration of their semantics, morphology, scope and relevance (Floridi, 2011). Philosophical questions are not only inspired by empirical and logic-mathematical resources, but also require other answers in the networks of philosophy as a science of open questions. The proliferation of big data has the potential to bring about significant changes to language cognition and natural language processing, thus opening up new research frontiers in philosophy and social science. This study examines the potential and necessity of a data turn in philosophy of language by reviewing the developmental context of modern philosophy. The purpose of this paper is to propose a data-driven approach to philosophy of language, which is both necessary and feasible in the epoch of big data, as well as to suggest future research directions in this area.

1 PHILOSOPHICAL AND COGNITIVE DILEMMA

1.1 DILEMMA OF COGNITIVE RESEARCH IN EARLY 20TH CENTURY

Western philosophy has followed an inherent theoretical logic trajectory that progressed from classical ontology research to modern epistemology research. Greek philosophers focused on ontology as their primary philosophical inquiry, aiming to generalize the world's ontological existence (Burnet, 1914; Preus; Anton, 1992). It is the unity, universality and eternity of all things that researchers use to comprehend and recognize diverse aspects of the world, thus gaining insight into the whole, essence, absoluteness and eternity of the universe. Over time, observational, experimental and mathematical research methods were gradually integrated into fields such as astronomy and mechanics, moving away from purely speculative methods. In contrast, philosophers commonly use natural language to express their ideas, with few conventions, norms or symbols.

Core issues in modern western philosophy encompass the origin, basis and standard of knowledge. In this epistemological turn, rationalism and empiricism are the two major schools of thought. Rationalism argued for the primacy of reason and deductive logic in

acquiring knowledge, whereas empiricism espoused the role of experience and inductive logic. Philosophers explored the problem of cognition from the perspectives of human existence and the world's ontology, resolving issues regarding the source of knowledge and the standard of truth. The result was an epistemological foundation for classical ontology and a move towards anti-theological thinking. Epistemology, however, faced a challenge of lacking cognitive tools to better understand the world. Philosophy stagnated due to a lack of language tools in the early twentieth century. As a result, mathematical logic emerged, which allowed philosophers to express their ideas in a symbolic and logical manner.

The rise of mathematical logic, born out of philosophical and cognitive dilemmas, has provided philosophers with tools to tackle language-related problems. Although philosophers at this stage are concerned with language and recognize its extreme importance, they prefer to completely transform natural language. They advocate that all philosophers, if not all people, should use clear, unambiguous and deducible mathematical logic to break down all complex ideas and things into simple propositions. Nevertheless, the artificially constructed language of mathematical logic could not entirely replace philosophical language and everyday language. Neither daily life nor cognitive research could switch completely to the symbolic language of mathematical logic. During the 1950s, analytical philosophy encountered a number of difficulties when attempting to reconstruct the world in a logical manner.

1.2 Emergence of language as an object in philosophy of language

As a distinctive feature of Western philosophy in the 20th century, language has assumed a pivotal role in philosophical inquiry (Rorty, 1967; Derrida, 1976, p. 527; Losonsky, 2006). The philosophy of language examines language in philosophical terms, exploring its essential nature, origin and usage as philosophical problems. Gottlob Frege's seminal paper *Sense and Reference* (1892), Bertrand Russell's seminal paper *On Denoting* (1905), Ludwig Wittgenstein's *Philosophical Investigations* (1953), John Austin's *How to Do Things with Words* (1962) and John Searle's *Speech Acts* (1969) are among the works credited with making a revolution to the development of philosophy of language.

Overlapping areas of research (e.g., theory of metaphor, translation, etc) enrich philosophy of language and enlarge its topics and issues. The idea of metaphor and metonymy by Jakobson (1956), which traces its origins to Saussure's dichotomy of horizontal and vertical combinations, is one of the most significant contributions to philosophy of language (Jakobson, 1956, p. 90). Language itself may need to undergo quite a substantial rethinking in light of a formal theory of communication. Using metaphors is not a matter of linguistic competency: rather, it is cognition at work. Interestingly, metaphoric cognition would show the universal features of human cognition and its fundamental independence of linguistic commitments.

In partial contrariety to this approach, the relevance theory by Sperber and Wilson (1986) advocates that utterances are linguistically encoded by systems of communicative relevance. Language understanding consists of possessing the sound interpretation for an utterance. Interpretation is nothing other than accessing the framework for representing the intention of speakers in the linguistic medium. As a consequence, cognition would depend on the particularity of natural language, and metaphorical cognition would work as any other case of linguistic act of thought. Lakoff and Johnson (1987, p. 73) revolutionize our understanding of language and how we relate to the world around us through conceptual metaphors. Metaphor is the act of interpreting and comparing one kind of thing with another. The metaphors we use in our everyday lives are fundamental to our brain's conceptual system. Lakoff and Johnson (1987, p. 73) approach metaphors from a conceptual viewpoint, with the aim of showing as metaphorical practices of language involve the possession of a web of concept (trans-contextually evident across different natural languages).

By constructing meaning through language, linguists (Halliday, 1978; Baker, 1996; House, 1997; Ong, 2010, p. 462) take a language-based approach to cognition. The concrete analysis of how different semantic systems of natural languages generate similar meanings is also at the ground of cultural studies, which, originating from Cassirer's seminal works (1996) on the transcendental symbolism of human cognitive apparatuses, include at present an irremovable concern for how languages work and represent the world across different traditions. The cultural rhetorical approach is proposed to provide a theoretical framework for explaining language as a communicative phenomenon related to the producer's and the receiver's cultural identities (Albaladejo, 2013, p. 1). Through the understanding of their rhetorical and literary use, the producer and the receiver can understand that these languages and the communication they facilitate are part of a special domain of cultural, linguistic and communicative conventions that differ from those of everyday life. Hatim and Mason (2014) describe the forms of language and translation knowledge required for translators to succeed. In their work, they illustrate how linguistics can be applied to translation creation, description and constructive criticism.

In the wake of the scientific revolution, philosophers shifted their focus from the nature of the knowable object to its epistemic relationship with the knowing subject, thereby moving from metaphysics to epistemology. A shift in philosophical focus to philosophy of logic and language has been prompted by the infosphere and the growth of information society. It is for this reason that the philosophy of information is concerned with the nature of information as well as its dynamics, such as communication, flow and processing (Floridi, 2011).

2 Paradigm shifts and new epistemologies in big data era

2.1 Data revolution and big data era

During the mid-20th century, electronic technology and the Internet enabled a major breakthrough in the acquisition, storage and computation of data. In the 21st century, social media and mobile devices have become more prevalent, as well as the explosion of data of various types. The term data traditionally refers to a combination of numbers and measurements used to describe things quantitatively, whereas big data is characterized by continuous production, fine-grained scope, flexibility and scalability (Kitchin, 2014, p. 1). In general, big data can be divided into two categories (Jin; Wah; Cheng *et al.* 2015, p. 59). The first type of data consists of observations and scientific experiments collected from the physical world, such as astronomical, biological and remote sensing data. A second type of data is data collected from human society, typically from the Internet, finance, economics and social network domains. A metaverse concept was proposed formally in 2021, suggesting that humans would soon live in a metaverse where the natural and data worlds are intertwined.

As compared to traditional data, big data is distinguished by the 5Vs: huge volume, high velocity, great variety, low veracity and high value (Jin; Wah; Cheng *et al.* 2015, p. 59). Massive data presents many challenges, not just its enormous volume. Real challenges arise from the variety of data types (variety), the speed with which the data should be processed (velocity) and the uncertainty associated with the data (veracity). It is important to note, however, that even the most effective methods of data cleansing, which eliminate some inherent unpredictability of the data, are insufficient to distinguish between true and false information, or between reliable and unreliable information. During the era of big data, data has evolved significantly in terms of its distinctive characteristics.

As a consequence of the technological revolution in data production and acquisition, the status of data has undergone significant changes. Wiener (1948) argues that the world is composed of three elements: matter, energy and information. The essence of data is the mapping or characterization of the world's state we live in (Dummett, 2014). The ability to perceive and understand information is an integral part of human intelligence. Data was previously seen as a supplement to other cognitive tools. Data represents language, characters, sounds, images and logical symbols, making it the most essential tool for representing information. Digitizing qualitative representations of the past enhances the status of data and gradually becomes its essence.

2.2 Paradigm Shifts and New Epistemologies

The indispensable and transformative role of data has been widely recognized in the era of big data and the metaverse (Olsher, 2014, p. 131; Agerri; Artola; Beloki *et al.*

2015, p. 36). Using heterogeneous data to answer real-world questions is the hallmark of data science (Rizk; Elragal, 2020, p. 1; Slota; Hoffman; Ribes *et al.* 2020, p. 1). As big data has revolutionized natural science research, humanities and social science research are faced with new opportunities. Through data prospecting, knowledge, expertise and practices are rendered available and tractable to data science methodologies and epistemologies. It consists of the upstream process of identifying disordered or inaccessible data resources, and rendering them available for computation by ordering and reorganizing them. While data prospecting is widely recognized for its practical benefits, methodological requirements have grown increasingly complex, necessitating philosophical guidance.

The digital revolution presents an essential opportunity for paradigm shifts and epistemological progress. With the advent of big data and new data analytics, established epistemologies in the sciences, social sciences and humanities have been challenged (Floridi, 2011). Additionally, it assesses the extent to which they shift paradigms across multiple fields, the philosophy of science and epistemology in particular (Symons; Alvarado, 2016, p. 1; Sætra, 2018, p. 508; Haig, 2020, p. 15; Costa, 2017, p. 27). Due to the diversity of philosophical underpinnings in the humanities and social sciences, the situation is more complex. According to Kitchin (2014, p. 1), the development of a situated, reflexive, contextually nuanced epistemology might be fruitful. In the foreseeable future, big data and new analytics will not establish radically different disciplinary paradigms. Big data reveals to be crucial for a variety of issues. The logical approach to language analysis may benefit from a quantitative empirical approach which widens the representational range of how concepts are used and understood. Metaphorical cognition and structural analysis of language, such as translation studies on meaning or cultural studies on the impact of languages on epistemic contexts, have today the opportunity of a data-driven research paradigm, which can foster an high-quality development of the philosophical understanding of language. As a result of technical methods, such as data mining, visualization, machine learning and virtual reality, philosophers can create novel research forms (Shardlow; Sellar; Rousell, 2022, p. 399; Kapchan, 1995, p. 479). For this reason, data openness, interoperability and comprehensive digital collaboration are essential for transforming philosophy research.

Data turns in philosophy are characterized by a shift from the physical world and human languages to data. In addition, the process involves the transition from speculative and analytical methods to synthesis, from formal logic and mathematical logic to algorithms, and from proof and causality to discovery and relevance. A logical, ontological and epistemological approach to data is presented here. Formal representation of knowledge tends to be more accurate with data than with language, that is, data can be mapped one-to-one with the world. Data language offers enhanced accuracy than natural language, and it can be calculated and modeled easier than logic language. Second, identifying the rules contained in the data can be done using calculations and algorithms. Data cleaning, classification, association,

aggregation and other processes enable us to discover rules and knowledge from fragments of knowledge, and to reintegrate and reconstruct them. Analytical methods tend to shift to synthesis-oriented methods. A large amount of data is used in order to discover a wide variety of useful knowledge that might not be absolutely reliable, but still has some utility. The cognitive goal and philosophical focus shift from knowledge proof to knowledge discovery. Big data has disrupted traditional scientific cognition that focuses on causality. According to Schönberger and Cukier (2013), our understanding of the world is further enhanced by non-causal analyses that emphasize what rather than why. Ultimately, scientific cognition shifts from a strict causality-based paradigm to a relevance-based paradigm.

3 RESPONSES TO DATA TURN OF PHILOSOPHY OF LANGUAGE

The emergence of big data has triggered significant multidimensional changes in philosophy of language, popularly referred to as the data turn (Gawde; Patil; Kumarm *et al.* 2023, p. 106). Various branches of philosophy are affected by big data, including political, moral, legal, scientific, and even aesthetic philosophy. As a result of the emergence of big data, philosophy of language has undergone significant multidimensional changes. It is just as pertinent to our times as we philosophize upon and about big data in this section.

3.1 CHANGE OF RESEARCH OBJECTS

Information is transmitted through language. Language research must address both human and machine needs in the modern era. This necessitates the conversion of natural language processing into language data (Agerri; Artola; Beloki *et al.* 2015, p. 36). Increasing emphasis is placed on the structure and content of information, as well as the use of scientific data formats to represent the information. The language of natural language pertains to the communication between humans, whereas the language of data represents the communication between humans and machines. Data language has thus paved the way for communication between humans and machines, which holds great promise for the future of technology.

As a result of the scientific cognition of data, the focus of philosophy of language has shifted to the realm of data, as researchers are now studying a variety of data-related issues, such as the nature of data, the relationship between data and the world, and the nature of algorithms. Such a change generates a number of debates among scholars. Controversies arose about the specific ontology of evidence at stake in big data research, particularly related to how to construe warranted beliefs and knowledge which may be evaluated soundly supported by collection of data. Two main options are at the theorist's disposal. The representational view understands data in terms of reliable interactions between humans and the world, any social setting being a structure able to generate dataization. In this light, data must pass test

of veracity (indicating the degree of reliability of a big data set - in absence of appropriate validation, the use of a set of data may lead to incorrect conclusion, Floridi; Illari, 2014; Cai; Zhu, 2015, p. 1) and validity (indicating the degree of appropriateness to a given use of a set of data - use of a set of data always requires explicit justification, Bogen, 2010, p. 778). On the contrary, the relational view thinks about data as theory related matters, namely, defines data in terms of evidence for scientific claims. Consequently, data do depend on their provenance, motivations and instruments used to visualize and use them being part of the support relation at work (Leonelli, 2016).

With data that has a one-to-one mapping relationship, it is now more accurate to describe the world through data. Consequently, scientific cognition promotes a strong philosophical work in the field of philosophy of data, with the aim of clarifying language analysis in terms of data transformation. Researchers now take data as the primary object of inquiry, with algorithms as their tools, and quantitative approaches as their methods. This combination of tools and methods can process both formal and natural language. In this context, philosophy of language draws upon the philosophy of data, information and intelligence to explore a wide range of topics, including data and the world, data and language, data and knowledge, data and truth, and data and ethics.

3.2 CHANGE OF RESEARCH METHODS

3.2.1 CONSTRUCTION OF SCIENTIFIC MODELS

Philosophers employ mathematical and computational models as fundamental tools for quantitative analysis. By using mathematical operations, researchers can analyze the relationship between objective phenomena and mathematical objects. Such models are supported by a substantial body of philosophical literature that helps scholars explain its nature and mechanism. As such, philosophers depend on algebra, differential equations and probability theory to tackle language-related issues and devise solutions accordingly.

Models derived from economics and evolutionary biology are used in philosophy of language. It is the interdisciplinary nature of philosophy of language that has led to accelerated progress in the field, correlating with the data turn of scientific cognition. Computational models are widely used and valuable in philosophy of language. These models represent the target system as data packets stored in algorithms and memory, allowing for the simulation of the subject of study based on the gathered data. By using iterative updates, researchers can collect data regarding the trajectory and final state of the system, allowing for further analysis and processing. Agent-based models are a common form of computer-based simulation used to understand interactions between semi-autonomous or autonomous agents. The purpose of

computer simulations, regardless of their form, is to aid researchers in understanding complex systems by augmenting mathematical models.

It offers several advantages to conduct language philosophy research using scientific models. Firstly, language is an outcome of both cultural and biological evolution, and it can be difficult to trace its evolutionary trajectory through lengthy prose packed with qualitative arguments. On the other hand, mathematical and computational models are better suited to the quantitative analysis of language because they provide interpretations where verbal reasoning falls short. The evolution of communication and rational decision-making would be difficult to trace without the use of mathematical and computational tools. The use of models can provide results that cannot be achieved through verbal reasoning, thereby increasing the reliability of philosophical research. Thus, computational and mathematical models make it possible for philosophers to gain a deeper understanding of phenomena that have traditionally been difficult to study. Using philosophy of language models, one can establish permissible and cut-off conditions, improving both accuracy and efficiency in processing vast and complex language phenomena.

3.2.2 APPLICATION OF DATA EXPERIMENT AND CORPUS ANALYSIS

Big data is characterized by an exponential growth in data volumes. Whether it is data mining, text processing, natural language processing, or the construction of machine models, most activities require a specific quantity of data. To construct models, researchers deploy rule-based methods or probability and statistics methods once data scale reaches a certain level. Throughout history, language philosophers have changed the way they analyze the function and structure of natural and artificial languages. Philosophy is now able to utilize big data platforms closely related to computer science, such as computational simulation experiments and corpus analysis (Machery, 2011, p. 191; Lau; Chan, 2022, p. 1). As the age of big data continues to evolve, data experiment analysis and corpus use are becoming more prevalent.

Philosophy of language benefits from corpus-based methods (Devitt, 2015; Spencer, 2020, p. 117). There is a fundamental problem with the assumption that more data would necessarily yield more information, namely, very large databases always contain arbitrary correlations. These correlations appear only due to the size, not the nature, of data. As a result, big data analysis is, by definition, unable to distinguish spurious from meaningful correlations and it is, therefore, a threat to scientific research (Calude; Longò, 2017, p. 595). Second, the use of complex software, in big data analysis, makes margins of error unknowable, because there is no clear way to test them statistically. The path complexity of programs with high conditionality imposes limits on standard error correction techniques (Symons; Alvaraddo, 2016, p. 1). Third, knowledge produced by analytic systems, as artificial intelligence, may be

at all unintelligible to humans. In case analysis is beyond the involved researcher's epistemic capability, knowledge derived from big data evidence may not involve an increase in human understanding, especially if understanding is understood as an epistemic skill (De Regt, 2017). A powerful reply to all these issues consists in invoking an improvement in technical mastery of computational methods, as statics and computer programming, by scholars in philosophy of language. In data-driven research, corpora provide observational data in comparison to experimental data. Research into corpus linguistics reveals that interpretations and results, based on corpus and experimental datasets, may diverge (Arppe; Jarvikivi, 2007, p. 131). This indicates that corpus research and experimental techniques should complement each other or drive new deeper discussions.

It is essential for language philosophers to consider corpus data in addition to experimental evidence and intuitive judgments (Erl; Khattak; Buller, 2016). Philosophical methodology should embrace both an armchair and a laboratory approach (Devitt, 2015). Corpus-based observational data that generalize authentic speech acts should also be utilized to construct conceptual theories of language. For certain philosophical problems, exploring a corpus of universal meaning may be challenging. In such situations, philosophers may be better suited to studying linguistic data produced in a laboratory setting. However, with the increasing information explosion and big data technologies, utilizing corpus data makes sense in many instances for addressing long-standing issues of interest to philosophers.

The use of corpus-based approaches holds enormous potential for addressing issues relating to language. Corpus research involves extracting, generalizing and analyzing the increasingly extensive language network. When it comes to phenomena, such as online language violence, underage negative online behavior, the proliferation of vulgar speech and inappropriate use of online buzzwords, prior studies have analyzed their semantic content, but they have often yielded one-sided and subjective conclusions. While scholars have made proposals and utilized philosophical analysis methods to evaluate these issues against actual language (Jiang; Bai, 2010, p. 3), certain questions still require resolution. For instance, are slurs always offensive if they pertain to the same object, and do they possess a consistent goal? How frequently are slurs appropriately contextualized? Despite these being questions of philosophical significance, they are better answered using corpus analysis. Corpus studies can precisely evaluate these questions by comparing them to actual language usage, providing increased reliability and accuracy to language philosophy research.

4 POTENTIAL SCOPE FOR DATA TURN OF PHILOSOPHY OF LANGUAGE

In the era of big data, every world's aspect can be represented by data, including words, sounds, images and languages (Lutz, 2012, p. 181). Similar to language, data serves as a universal means of expressing oneself. Scientific research seeks to analyze various

developmental situations and explore the underlying connections among these forms, relying on sufficient and fact-based data resources. By leveraging quantitative analysis in philosophy of language, various contexts, developmental patterns of issues and human behavior can be summarized or predicted. Incorporating different perspectives, in relation to big data, would be fruitful. Clashes have been identified between assumptions underlying computational approaches to social data and the context of their production (Boyd; Crawford, 2012, p. 662; Törnberg; Törnberg, 2018, p. 1). Due to a lack of philosophical discussion on social reality in the digital age, big data may pose ethical dilemmas for different branches of philosophy (Mittelstadt, 2019, p. 17). Törnberg and Törnberg (2018, p. 1) propose a metatheoretical perspective and a stable ontological position, which enable us to hear message of the data about the social world and listen what they fail to convey. Basden and Klein (2008, p. 260) point out new research directions for data and knowledge engineering. The future of data mining and natural language processing can address a wider range of meanings than can be inferred from analyzing text using current methods that focus on syntactic and semantic meanings. The following aspects summarize the potential scope for data turn in philosophy of language.

4.1 Transforming qualitative and quantitative research paradigms

For over half a century, philosophers of language, as well as other disciplines, have debated the compatibility between qualitative concepts and quantitative tools. Ayer (1946) argues that philosophy should solely focus on language and logic during the height of logical positivism. Others (Machery, 2011, p. 191), however, strongly oppose this viewpoint because it largely ignores the importance of quantitative research methods. Qualitative and quantitative analysis are distinct research approaches. The evolution of language meaning in philosophy of language, for example, is better studied quantitatively with computer simulations and dynamic models. In contrast, qualitative analysis is more appropriate when the researcher is able to observe directly linguistic patterns or phenomena without the influence of sampling bias. Research within the context of big data has transformed quantitative and qualitative research paradigms and fostered an intersectional zone. The vast scale and characteristics of big data have made it possible for quantitative and qualitative research techniques to gradually converge in terms of data acquisition and analysis. To some extent, this has alleviated or even reconstructed the relationship between quantitative and qualitative research. The existence of more data does not imply the resolution of all problems. With big data, comes a wider range of data that can be analyzed, and novel techniques and approaches can be developed. It will complement, but never replace, small data studies. On the other hand, big data analysis advancement can be compatible with qualitative research (King; Keohane; Verba, 2021).

4.2 Reconstructing quantitative approach to philosophy of language

Quantitative approaches to philosophy of language utilize quantitative methods in order to understand how language shapes our relationship with the world. In addition to providing an unprecedented opportunity to consolidate data, big data can enhance the completeness of data to unprecedented levels (Olsher, 2014, p. 131; Westhaver, 2021, p. 161). Concepts are presented by concrete linguistic expressions, and the words used to express them reflect spatial-temporal variation of conceptual properties. Using quantitative analysis methods, researchers can examine the diachronic and synchronic dimensions of the philosophy of language in order to optimize measurement variables and trigger unexpected results. Quantitative analysis of this aspect requires searching across the semantic and intellectual web for relevant properties and their datum contexts. The quantitative analysis of conceptual changes, in the era of big data, is reflected in intra-lingual and inter-lingual contexts. With the expansion of the observation range, quantitative measurement methods may be reused in secondary analyses of existing research. It is not only the access to data that expands the observation range, but also the analysis of data using advanced computer tools, particularly artificial intelligence. Artificial intelligence has been used, in a number of ways, to facilitate the collection, structuring and analysis of big data (O'Leary, 2013, p. 96). However, current technologies still experience difficulties in handling all real-time scenarios for data analysis. There is ample potential for research to address these challenges, providing new dimensions for interdisciplinary research in this field.

4.3 Expanding human-data relation in philosophy of big data

Data sets with a high volume, variety, velocity, veracity and variability have emerged as one of the most significant developments of the past decade. The philosophy of big data examines data sets at two levels. An internal discussion focused on conceptualization, knowledge possibilities and truth standards, while an external discussion focused on the implications of big data for individuals, society and global issues. In addition to automating mechanical tasks, big data science and services can also automate cognitive tasks, reshaping the human-data relationship and enabling subjectivation as well as fulfilling emotional and intellectual needs. Through data visualization technologies, information can be represented, reproducing knowledge through graphics, time series, layers and other elements that help visualize the issue in question (Sun; Li; Jiang, 2019, p. 25). Meaning can be extracted from data when patterns, trends and dependencies are visualized. Through a deliberative systems approach, data visualization can be integrated into the overall subjectivation and normatively assessed in terms of key conceptualization and truth standards (O'Leary, 2013, p. 96). An increasing demand exists for data visualization that provides insights into individuals, society and the world at large. Humans and machines have been brought together in concrete

practices that align human, technological and linguistic dimensions. There is great potential for fostering mutual growth and collaboration between humans and data entities through the philosophy of big data.

Conclusions

This paper reviews the data turn of philosophy of language in the epoch of big data. Revisiting the development of philosophical thoughts from ontological to epistemological patterns, it describes theoretical interests of the philosophy of language in transforming qualitative and quantitative research paradigms. A situated, reflexive, contextually nuanced epistemology is proposed. Data turn involves the transition from speculative and analytical methods to synthesis, from formal logic to algorithms, and from proof to relevance and discovery. The emergence of big data has triggered changes in language philosophy research objects and methods. Accordingly, language philosophers have explored a variety of data-related issues, including the nature of data, the relationship between data and the world, and the nature of algorithms. Philosophy of language powered by data provides humanity with a way to gain a deeper understanding of the world and build it responsibly. Philosophy research, in this area, is expected to become one of the most exciting and worthwhile in the near future.

References

AGERRI, R.; ARTOLA, X.; BELOKI, Z. *et al.* Big Data for Natural Language Processing: A Streaming Approach. **Knowledge-Based Systems**, v. 79, n. 5, p. 36-42, 2015.

ALBALADEJO, T. Retórica Cultural, Lenguaje Retórico Y Lenguaje Literario. **Tonos Digital**, v. 25, p. 1-21, 2013.

ARPPE, A.; JARVIKIVI, J. Every Method Counts: Combining Corpus-based and Experimental Evidence in the Study of Synonymy. **Corpus Linguistics and Linguistic Theory**, v. 3, n. 2, p. 131-159, 2007.

AUSTIN, J. L. **How to Do Things with Words**. Oxford: The Oxford University Press, 1962.

AYER, A. J. **Language, Truth and Logic**. London: Victor Gollancz, 1946.

BAKER, M. C. **The Polysynthesis Parameter**. Oxford: Oxford University Press, 1996.

BASDEN, A.; KLEIN, H. K. New Research Directions for Data and Knowledge Engineering: A Philosophy of Language Approach. **Data & Knowledge Engineering**, v. 67, n. 2, p. 260-285, 2008.

BOGEN, J. Noise in the World. **Philosophy of Science**, v. 77, n. 5, p. 778-791, 2010.

BOYD, D.; CRAWFORD, K. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. **Information, Communication & Society**, v. 15, n. 5, p. 662-679, 2012.

BURNET, J. **Greek Philosophy**: Thales to Plato. London: Macmillan, 1914.

CAI, L.; ZHU, Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. **Data Science Journal**, v. 14, n. 2, p. 1-10, 2015.

CALUDE, C.S.; LONGÒ, G. The Deluge of Spurious Correlations in Big Data. **Foundations of Science**, v. 22, n. 3, p. 595-612, 2017.

CASSIRER, E. **The Philosophy of Symbolic Forms**: Volume 4 - The Metaphysics of Symbolic Forms. New Haven: Yale University Press, 1996.

COSTA, A. da. The Pardoner's Passing and How It Matters Gender, Relics and Speech Acts. **Critical Survey** v. 29, n. 3, p. 27-47, 2017.

DE REGT, H. W. **Understanding Scientific Understanding**. Oxford: Oxford University Press, 2017.

DERRIDA, J. The Supplement of Copula: Philosophy before Linguistics. **The Georgia Review**, v. 30, n. 3, p. 527-564, 1976.

DEVITT, M. Testing Theories of Reference. *In*: HAUKIOJA, J. (ed.). **Advances in Experimental Philosophy of Language**. London & New York: Bloomsburry, 2015.

DUMMETT, M. **Origins of Analytical Philosophy**. Bloomsbury: Bloomsbury Academic, 2014.

ERL, T.; KHATTAK, W.; BULLER, P. **Big Data Fundamentals**: Concepts, Drivers and Techniques. Boston: Prentice Hall, 2016.

FLORIDI, L.; ILLARI, P. **The Philosophy of Information Quality**. Cham: Springer International, 2014.

FLORIDI, L. **The Philosophy of Information**. Oxford: Oxford University Press, 2011.

FREGE, G. **Über** Sinn und Bedeutung. **Zeitschrift für Philosophie und philosophische Kritik**, v. 100, p. 25-50, 1892.

FURNER, J. Philosophy of Data: Why? **Education for Information**, v. 33, n. 1, p. 55-70, 2017.

GAWDE, S.; PATIL, S.; KUMAR, S. *et al*. Multi-fault diagnosis of Industrial Rotating Machines using Data-driven approach: A review of two decades of research. **Engineering Applications of Artificial Intelligence**. v. 123, Part A, August 2023, p. 106-139, 2023.

HAIG, B. D. Big data science: A Philosophy of Science Perspective. *In*: WOO, S. E.; TAY, L.; PROCTOR, R. W. (ed.). **Big Data in Psychological Research**. Washington: American Psychological Association, p. 15-33, 2020.

HALLIDAY, M. A. K. **Language as Social Semiotic**. London: Edward Arnold, 1978.

HATIM, B.; MASON, I. **Discourse and the Translator**. New York: Routledge, 2014.

HOUSE, J. **Translation Quality Assessment**: A Model Revisited. Tübingen: Gunter Narr, 1997.

JAKOBSON, R. The Metaphoric and Metonymic Poles. *In*: JAKOBSON, R.; HALLE, M. (ed.). **Fundamentals of Language**. The Hague/Paris: Mouton, p. 90-96, 1956.

JIANG, Y.; BAI, T. Studies in Analytic Philosophy in China. **Synthese**, v. 175, n. 1, p. 3-12, 2010.

JI, C.; LI, Y.; QIU, W. *et al*. Big Data Processing: Big Challenges and Opportunities. **Journal of Interconnection Networks**, v. 13, n. 3, p. 1-19, 2012.

JIN, X.; WAH, B. W.; CHENG, X. *et al*. Significance and Challenges of Big Data Research. **Big Data Research**, v. 2, n. 2, p. 59-64, 2015.

KAPCHAN, D. A. Performance + In folklore. **Journal of American Folklore**, v. 108, n. 430, p. 479-508, 1995.

KING, G.; KEOHANE, R. O.; VERBA, S. **Designing Social Inquiry**: Scientific Inference in Qualitative Research. Princeton: Princeton University Press, 2021.

KITCHIN, R. Big Data, New Epistemologies and Paradigm Shifts. **Big Data & Society**, v. 1, n. 1, p. 1-12, 2014.

LAKOFF, G.; JOHNSON, M. The Metaphorical Logic of Rape. **Metaphor and Symbol**, v. 2, n. 1, p. 73-79, 1987.

LAU, J. Y. F.; CHAN, J. K. L. A Brief History of Analytic Philosophy in Hong Kong. **Asian Journal of Philosophy**, v. 1, n. 1, p. 1-20, 2022.

LEONELLI, S. **Data-Centric Biology:** A Philosophical Study. Chicago: University of Chicago Press, 2016.

LOSONSKY, M. **Linguistic Turns in Modern Philosophy**. Cambridge: Cambridge University Press, 2006.

LUTZ, S. Artificial Language Philosophy of Science. **European Journal for Philosophy of Science**, v. 2, n. 2, p. 181-203, 2012.

MACHERY, E. Thought Experiments and Philosophical Knowledge. **Metaphilosophy**, v. 42, n. 3, p. 191-214, 2011.

MITTELSTADT, B. The Ethics of Biomedical 'Big Data' Analytics. **Philosophy & Technology**, v. 32, p. 17-21, 2019.

O'LEARY, D. E. Artificial Intelligence and Big Data. **IEEE Intelligent Systems**, v. 28, n. 2, p. 96-99, 2013.

OLSHER, D. Semantically-based Priors and Nuanced Knowledge Core for Big Data, Social AI, and Language Understanding. **Neural Networks**, v. 58, n. 10, p. 131-147, 2014.

ONG, Y. P. The Language of Advertising and the Novel: Naipaul's A House for Mr. Biswas, **Twentieth Century Literature** v. 56, n. 4, p. 462-492, 2010.

PREUS, A.; ANTON, J. P. **Essays in Ancient Greek Philosophy V**: Aristotle's Ontology. New York: State University of New York Press, 1992.

RIZK, A.; ELRAGAL, A. Data Science: Developing Theoretical Contributions in Information Systems Via Text Analytics. **Journal of Big Data**, v. 7, p. 1-26, 2020.

RORTY, R. M. **The Linguistic Turn**: Recent Essays in Philosophical Method. Chicago: University of Chicago Press, 1967.

RUSSELL, B. On denoting. **Mind**, v. 14, n. 4, p. 479-493, 1905.

SÆTRA, H. S. Science as a Vocation in the Era of Big Data: The Philosophy of Science behind Big Data and Humanity's Continued Part in Science. **Integrative Psychological and Behavioral Science**, v. 52, n. 4, p. 508-522, 2018.

SCHÖNBERGER, V. M.; CUKIER, K. **Big Data**: A Revolution that will Transform How We Live, Work and Think. New York: Houghton Mifflin Harcourt, 2013.

SEARLE, J. **Speech Acts.** An Essay in the Philosophy of Language. Cambridge: Cambridge University Press, 1969.

SHARDLOW, M.; SELLAR, S.; ROUSELL, D. Collaborative Augmentation and Simplification of Text (CoAST): Pedagogical Applications of Natural Language Processing in Digital Learning Environments. **Learning Environment Research**, v. 25, n. 2, p. 399-421, 2022.

SLOTA, S. C.; HOFFMAN, A. S.; RIBES, D. *et al*. Prospecting (in) the Data Sciences. **Big Data & Society**, v. 7, n. 1, p. 1-12, 2020.

SPENCER, M. Pali Grammar: The Language of the Canonical Texts of Theravada Buddhism, vol 1 **Buddhist Studies Review**, v. 37, n. 1, p. 117-126, 2020.

SPERBER, D.; WILSON, D. **Relevance**: Communication and Cognition. Cambridge, MA: Harvard University Press, 1986.

SUN, G.; LI, F.; JIANG, W. Brief Talk about Big Data Graph Analysis and Visualization. **Journal on Big Data**, v. 1, n. 1, p. 25-26, 2019.

SYMONS, J.; ALVARADO, R. Can We Trust Big Data? Applying Philosophy of Science to Software. **Big Data & Society**, v. 3, n. 2, p. 1-17, 2016.

TÖRNBERG, P.; TÖRNBERG, A. The Limits of Computation: A Philosophical Critique of Contemporary Big Data Research. **Big Data & Society**, v. 5, n. 2, p. 1-12, 2018.

WESTHAVER, G. Continuity and Development: Looking for Typological Treasure with William Jones of Nayland and E. B. Pusey. **Bulletin of the John Rylands Library**, v. 97, n 1, p. 161-177, 2021.

WIENER, N. **Cybernetics, or Communication and Control in the Animal and the Machine**. Cambridge, MA: MIT Press, 1948.

WITTGENSTEIN, L. **Philosophical Investigations**. G.E.M. Anscombe and R. Rhees (ed.), G.E.M. Anscombe (trans.). Oxford: Blackwell, 1953.