

Article

Quantum Natural Language Processing: Challenges and Opportunities

Raffaele Guarasci, Giuseppe De Pietro  and Massimo Esposito * 

Institute for High Performance Computing and Networking (ICAR), National Research Council (CNR), 80131 Naples, Italy; raffaele.guarasci@icar.cnr.it (R.G.); giuseppe.depietro@icar.cnr.it (G.D.P.)

* Correspondence: massimo.esposito@icar.cnr.it

Abstract: The meeting between Natural Language Processing (NLP) and Quantum Computing has been very successful in recent years, leading to the development of several approaches of the so-called Quantum Natural Language Processing (QNLP). This is a hybrid field in which the potential of quantum mechanics is exploited and applied to critical aspects of language processing, involving different NLP tasks. Approaches developed so far span from those that demonstrate the quantum advantage only at the theoretical level to the ones implementing algorithms on quantum hardware. This paper aims to list the approaches developed so far, categorizing them by type, i.e., theoretical work and those implemented on classical or quantum hardware; by task, i.e., general purpose such as syntax-semantic representation or specific NLP tasks, like sentiment analysis or question answering; and by the resource used in the evaluation phase, i.e., whether a benchmark dataset or a custom one has been used. The advantages offered by QNLP are discussed, both in terms of performance and methodology, and some considerations about the possible usage QNLP approaches in the place of state-of-the-art deep learning-based ones are given.

Keywords: quantum computing; natural language processing



Citation: Guarasci, Raffaele, Giuseppe De Pietro, and Massimo Esposito. Quantum Natural Language Processing: Challenges and Opportunities. *Appl. Sci.* **2022**, *12*, 5651. <https://doi.org/10.3390/app12115651>

Academic Editor: Julian Szymanski, Higinio Mora, Doina and Andrzej Sobiecki

Received: 22 March 2022

Accepted: 1 June 2022

Published: 2 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, the explosion of neural language models based on deep learning architecture has led to significant improvement in all NLP tasks [1–4], ranging from machine translation [5], text classification [6], coreference resolution [7,8] or multi-language syntactic analysis [9–11]. In particular, Transformers-based models such as BERT have proved to outperform previous generation state-of-the-art architecture such as Long Short-Term Memory (LSTM) recurrent neural networks (RNN).

However, the improvement in performance is matched by an increasing complexity of models that have led to a paradox. Models require a huge amount of data to be efficiently trained, with an enormous cost in time, resources and computation.

This is the major drawback of current approaches based on Transformers, for instance the number of parameters for this kind of neural networks reaches the order of hundreds of billions (data referred to OpenAI GPT model) [12,13]. In addition, it requires big resources for the training phase (e.g., the whole Wikipedia corpus in several languages).

Beyond to these aspects, there are also open issues inherent to what really these models learn about language [14,15], how they encode this information [16] and how much of the information learned is really interpretable [17]. The literature has produced several studies focused on whether neural language models are able to encode a sort of linguistic information or whether they just replicate patterns observed in written texts.

An alternative way that is gaining attention in recent years is that which originates from quantum computing, in particular quantum-machine learning sub-field. The idea is to exploit powerful aspects borrowed from quantum mechanics to overcome computational limitations of current approaches [18]. The dominant paradigm of classical statistics

could be extended using quantum mechanics by representing objects with matrices of complex numbers.

In quantum computing, bits are replaced by qubits, which are able to handle information in a non-binary state using a property of quantum system called superposition [19]. Quantum algorithms can perform calculations with smaller complexity compared to the classical approaches using an intrinsic property of qubits known as super-polynomial speedup [18,20,21].

Recent work has observed the development of quantum algorithms that could serve as foundations for machine learning applications. In some cases, quantum properties have been used to simply improve performance of machine learning approaches, while in others problems of machine learning have been reformulated using quantum theory [22,23]. Process classical data using machine learning algorithms using quantum systems has generated a big strand of research. Quantum machine learning has been used for different purposes: delving into the use of quantum phenomena for learning systems, exploring the ability of quantum computers to learn on quantum data and the possibility to reformulate and implement machine learning algorithms on quantum hardware.

As happened before in the case of the classic machine learning, this rapid growth of quantum machine learning algorithms, both from hardware and software side (in terms of devices and algorithms) has involved the natural language processing (NLP) field. This has given rise to the so-called QuantumNLP (QNLP) [24], defined as the implementation of natural language on quantum hardware. It matches the compositional language structure (grammar and semantics) with composition made possible by quantum systems in order to model natural language and perform simple NLP tasks exploiting algorithms derived from the quantum machine learning.

The core idea at the basis of QNLP is that the most effective manner for bringing language meaning and grammatical structure together was provided by the categorical quantum mechanics framework. Moreover, it promises to be more aware of the meanings of the language involved in the modeling. The advantage of using a visual representation language is the possibility to represent both the meaning of a word and its relationships to others within a sentence, and its structure too. This would represent a step forward from the classical dependency-based representation that relies on tree structure without explicit connections to meanings of the syntactic elements involved.

The theoretical framework common to *de facto* all QNLP approaches is the Categorical Distributional Compositional (DisCoCat) model for natural language [25]. DisCoCat allows encoding meanings of words and phrases as quantum states and processes subsequently implementable as quantum circuits in dedicated hardware or simulators. At the moment, the developed approaches can mainly be pursued through a hybrid methodology, combining classical operations with operations taken from quantum mechanics. This is because that—although the advantages of quantum-based approaches are proven at the theoretical level [25,26]—the availability of hardware is currently limited to machines only to small to intermediate scale (in particular NISQ computers: Noisy intermediate-scale) [27]. Given the growing interest of the NLP community in the possibilities offered by QNLP, this paper aims at providing a comprehensive overview of quantum approaches to solve natural language processing tasks. After a brief introduction of concepts and properties drawn from quantum mechanics and NLP, the theoretical background underlying the proposed approaches is described.

Subsequently, different work are listed distinguishing between their type, i.e., full theoretical work that have not had a real implementation, work based on quantum mechanics but running on classical hardware, and work that can run on quantum hardware currently available and have been tested on real data.

Note that—although QNLP is promising and could solve many critical issues related to current well-established NLP mechanics—it is not yet possible to make a real comparison between classical and quantum approaches [28]. This is since many of the works have only

theoretically demonstrated performance benefits, while those that have been implemented only manage to run on very small portions of data.

This paper is structured as follows: Section 2 introduces the theoretical background both from a quantum computing and linguistic point of view. First, in Section 2.1 basic quantum mechanics notions, algorithms, and hardware used within the QNLP are specified. Section 2.2 briefly describes the intersection between quantum computing and NLP. After that, in Section 3 the approaches of QNLP that have been proposed so far are enumerated, dividing them into theoretical (Section 4), classical hardware approaches (Section 5) and approaches running on real quantum hardware (Section 6). Sections 7 and 8 are devoted to a review of the listed work and conclusions highlighting the advantages and disadvantages of QNLP approaches.

2. Theoretical Background

This section aims to provide preliminary concepts with respect to both the terminology, properties and algorithms derived from quantum mechanics that are discussed below and concerning the linguistic theory foundations that underlie the proposed language models.

2.1. Quantum Background

In this section fundamental quantum concepts that are used in this work are briefly outlined, covering both the fundamental concepts and the algorithms and hardware used within the QNLP.

2.1.1. Fundamental Notions

- **Qubits** is the basic unit of information in quantum computing. Similar to its classical counterpart, the bit, it can assume two distinct values of 0 or a 1. The difference is that whereas a bit must be either 0 or 1, a qubit can be 0, 1 or a superposition of both. Conventionally, possible states of a qubit are represented using the Dirac notation: $|0\rangle$ and $|1\rangle$.
- **Superposition** is a principle of quantum mechanics allowing separate elements to assume many configuration arrangements, such that the general state is a combination of all of these possibilities. For the purpose of this work, notice that each qubit could take a superposition of both $|0\rangle$ and $|1\rangle$, assuming a representation obtained from the linear combination of the two states: $|\psi\rangle = a|0\rangle + b|1\rangle$, where $|\psi\rangle$ is a state that lives in a 2-dimensional complex vector space (Hilbert space), a and b two arbitrary complex coefficients whose sum of squares is 1 and $|0\rangle$ and $|1\rangle$ are orthonormal basis vectors, related to the respective measures 0 and 1. In the context of NLP, the superposition gives the possibility to better manage some pervasive natural language phenomena, such as lexical ambiguity and polysemy [29,30]. For instance, the term *bar* can have different meanings. It can indicate a place where alcoholic drinks are served or a piece of metal. Using the Dirac notation, the word *bar* can be represented as a superposition state: $|\bar{a}\rangle = a|\text{place}\rangle + b|\text{rod}\rangle$.
- **Entanglement** is a non-local property that distinguishes qubits from the classic bit. It allows multiple states simultaneously, differently from the classical bits that can have only one value at a time. For instance, considering two entangled qubits in the Bell state, Bell states are specific quantum states of two qubits representing the simplest and maximal examples of quantum entanglement $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. These two qubits are separated and assigned to two different elements, a and b . The measurement of qubits assigned to a gives $|0\rangle$ or $|1\rangle$ as a result. Due to the entanglement property, b must now obtain exactly the same measurement as a . In particular, if a measures a $|0\rangle$, b must measure the same, as $|00\rangle$ is the only state where a qubit is a $|0\rangle$. With regard to NLP, the assumption is that words that the grammatical structure that connects with specific relations of multiple words, form and entangle between quantum states in which the words are encoded [31].

- **Hilbert space** is defined as a vector space equipped with an inner product, which allows defining a distance function so that it becomes a complete metric space. In quantum mechanics, the possible states of a system are represented by state vectors located in a complex separable Hilbert space [32].
- **Quantum measurement** is the act of observing a qubit in superposition and resulting in one of the possible states. The act of observing or measuring a qubit collapses the superposition state and the qubit takes on a classical binary state of either 1 or 0. Thus, when a qubit in a certain superposition state $|\psi\rangle = a|0\rangle + b|1\rangle$ is measured with respect to the standard basis for quantum computation $|0\rangle, |1\rangle$, it is possible to get either the result 0 with probability a^2 , or the result 1 with probability b^2 . Once a qubit has been measured, it stays in that state forever. In other words, measurement alters the state of a qubit, collapsing it from its superposition of $|0\rangle$ and $|1\rangle$ to the specific state consistent with the result of the measurement, i.e., if $|\psi\rangle$ is observed to be in state $|0\rangle$ through measurement, then the post-measurement state of the qubit will be $|0\rangle$, and any subsequent measurements (in the same basis) will yield 0 with probability 1.
- **Quantum interference** is a principles of quantum theory that affects the state of a qubit to influence its probability to collapse into a manner or another. It allows biasing the measurement of superposition toward a desired basis state or set of states.

2.1.2. Quantum Computation and Algorithms

- **Noisy intermediate-scale quantum (NISQ) devices** are defined as quantum computers for which general-purpose quantum error correction is not feasible and hardware errors are expected [33]. These devices currently can handle 1000 two-qubit operations with tolerable error rates, having a memory size of 50–100 qubits. NISQ devices are currently the only hardware in place to run quantum algorithms; however, they present several limits concerning the number of qubits available to algorithms and the maximum size of quantum circuits.
- **Quantum Random Access Memory (QRAM)** is the quantum counterpart of the classical random access memory (RAM). QRAM is able to use n qubits to address any quantum superposition of N memory cells, where the classical RAM can only use n bits to randomly access $N = 2^n$ distinct memory cells. QRAM presents some advantages in terms of performance, exponentially reducing the need for a memory call from N to $O(\log N)$. The architecture has been presented by [34], but it still remains unrealized at the implementation level.
- **Grover's quantum search algorithm** is structured to get the correct answer using a state with an equal superposition of orthogonal states, which represent the answers as input. The output is a state with only basic states that have any probability of being measured correspond to correct answers [35].
- **Quantum volume** is the metric that allow a comparison between different quantum computers, each of which can present a different architecture. Performances are evaluated quantifying the largest random circuit of same width and depth successfully implementable by a quantum computer.

2.2. NLP and Quantum: The Meeting Point

One of the assumptions underlying the union between natural language processing and quantum theory is the possibility of creating a direct relationship between linguistic features (i.e., syntactic structures and semantics meanings) and quantum states.

This is made possible using the DisCoCat framework through string diagrams [36] as a network-like language [37].

This approach is part of a long and flourishing tradition of computational linguistics focused on the search for the most efficient way to represent language structures and meanings in a machine-readable way. On the one hand, the distributional approach—which has been the most successful line of research in recent years—relies on statistics about the contexts in which words occur according to the distributional hypothesis [38]. By contrast,

the symbolic approach [39] has been focused on individual meanings that compose the sentence. This approach is based on the theoretical linguistics’ concept of compositionality, arguing that the meaning of a sentence depends on the meanings of its parts and by the grammar according to which they are arranged together. Therefore, the analysis of the individual constituents determines the overall meaning, which is expressed using a formal logical language. This line of research has obtained less success in NLP applications so far.

Current state-of-the-art neural network models are based on the dominant distributional paradigm. Therefore, this approach is not without problems. First, there is a big bottleneck created by the need for ever larger data sets and parameters; moreover, the interpretation of these models is difficult [40].

The first attempt to overcome the limitations of current NLP models is to include features about the structure of the language (basically syntax) into canonical distributional language models. The resulting model—denoted as DisCoCat—incorporates categorical information and distributional information. Note that this is certainly not a new approach in the field of theoretical and computational linguistics, since its roots lie in the Universal Grammar [41] and foundational work of [42,43], while applied aspects come from categorical grammars proposed by [44] and pregroup grammar [45].

The Compositional Distributional Model

Given the premise that the constituents of a sentence are strongly interconnected, and the grammatical structures in which they are involved affect semantics [46], the pioneering work proposed by [25] has proposed a graphical framework to draw string diagrams (see Figure 1) exploiting concepts from Lambek’s pregroup grammar [47]. The uniqueness of the proposed representation is that sentence meanings can be totally independent of the grammatical structure.

The question they intended to answer is not only rooted in compositionality, i.e., whether the meaning of a whole sentence can be deduced by single meanings of its words. The aim is rather to make the first steps towards a grammar-informed NLP, deepening the ways in which words interact with each other and establishing their meanings. In other terms, the framework aims to combine in a whole diagrammatic representation structural aspects of language (grammar theory and syntax) and statistical approaches based on empirical evidences (machine/deep learning).

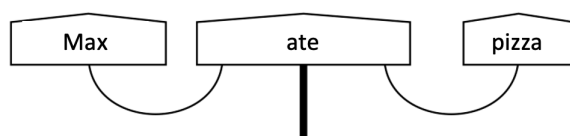


Figure 1. Example of a simple sentence represented using a string diagram inspired by formalism proposed in [25].

In the diagram, boxes represent meanings of words that are transmitted via wires. It deals with a representation similar to the canonical Dependency Parse Tree (DPT) well known in the linguistics literature, but it does not introduce a hierarchical tree structure. In the example shown in Figure 1, the noun in subject position “Max” and the one in object position “pizza” are both related with the verb “ate” and the combination of these words builds up the meaning of the overall sentence. In this way, distributional and compositional aspects are combined into DisCoCat. The meaning of sentences is computed using pregroup grammar via tensor product composition. In particular, it is possible to go through the classic DPT using the tensor product of vector spaces of the meanings of words and vectors of their grammatical roles. For instance, the example sentence in Figure 1 can be represented as follows:

$$(\overrightarrow{Max} \otimes \overrightarrow{subj} \otimes \overrightarrow{ate} \otimes (\overrightarrow{pizza} \otimes \overrightarrow{obj})) \tag{1}$$

This vector in the tensor product space can be considered as the meaning of the sentence “Max ate pizza”. Subsequently, this model has been reformulated in quantum terms, creating the pregroup only using Bell-effect and identities [24]. In this diagrammatic notation (see Figure 2), pentagons represent quantum states and wires represent the Bell-effect. The equivalence of wire structure with pregroup grammar has been demonstrated [31].

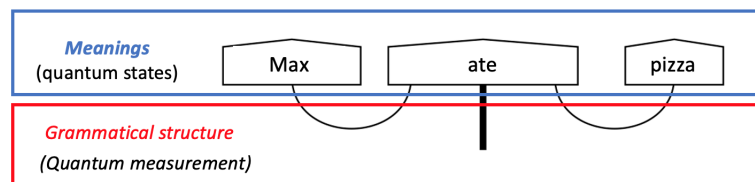


Figure 2. Diagrammatic notation showing how word meaning can be interpreted as quantum states and grammatical structure as quantum measurements.

Notice that the original DisCoCat model works perfectly without any reference to quantum theory, even if its true origin is the categorical quantum mechanics (CQM) formalism [48] and this connection is only made explicit in further work [31].

The novelty in introducing elements from quantum theory lies in the argument put forward in the work of [49] and then elaborated and enriched in [31]: QNLP can be considered “quantum-native” since quantum theory and natural language share an interaction structure and the use of vector spaces. This interaction structure determines the entire structure of processes, including the specification of the spaces where the states live. Vector spaces are used to describe states. This implies that natural language could better fit in a quantum hardware than a classical one.

Hence, the translation of linguistic structure into quantum circuits is particularly suitable to be implemented in a proper quantum hardware (NISQ) and consequently benefit from quantum advantage in terms of speedup.

3. Classification of Work

The release of a compositional derivational model such as DisCoCat, has tremendously influenced in direct and indirect ways all approaches in the nascent field of QNLP. Since in its original version it can actually be considered as a tensor network language model [50], some approaches have taken advantage of its internal architecture while not using the whole model. Other works instead incorporate DisCoCat into their methodology, while still other ones propose alternative models.

The papers reviewed below are categorized according to this criterion. First, a brief mention of pre-quantum work is given, i.e., all work using the first version of the model or work exploiting tensor networks to perform different NLP tasks. Although such works may not appear to be in line with the purpose of this survey, it is important to include them since they lay the groundwork for all future developments. In addition, they cover a broader range of topics from which subsequent work complete with a full quantum implementation will branch.

After that, Theoretical Quantum work are described, i.e., those works that propose algorithms or methods based on quantum theory or are potentially implementable on quantum, hardware, but which are not tested on real data. They range from approaches focused on specific tasks to other ones that targeted generic aspects of NLP, such as how to represent sentences or how to combine different language models (e.g., syntactic/semantic or compositional/distributional).

Then, approaches that have been implemented and evaluated on real data are described. A distinction is made between approaches that have been run on classical hardware (namely quantum-inspired approaches) and those that have used actual quantum hardware.

Concerning quantum-inspired approaches, the quantum advantage has been demonstrated at the theoretical level but the algorithms have been tested on classic hardware.

These works are typically organized in this way: taking up the classical model, benefiting from the contribution of quantum mechanics at the performance level and demonstrating that it is possible to outperform the state-of-the-art using a quantum approach. Note that some of these works have been tested and compared with benchmark datasets well-known in the literature, while other ones use ad-hoc-created samples for the evaluation.

Finally, approaches that have been actually tested on the quantum computer are outlined. These works—although at the moment very few in number—have been proposed for some simple NLP tasks. These works are described following a distinction with respect to the application task. These work are intended as the applied counterpart of theoretical works in which the mathematical foundations are provided. All of them start from the assumption that a quantum-based model of language should be closer and more reliable than current language models with respect to a specific task. Experiments have been performed on small-to-medium scale datasets, due to the current limits of quantum hardware.

Table 1 shows a comparative summary of the listed approaches, specifying type and task.

Table 1. Comparison of existing approaches divided by type (theoretical, really implemented on classical hardware and running on quantum hardware) and by NLP task on which they focused.

Type	Task	Work
Theoretical approaches	Syntactic-semantic Representation	Zeng and Coecke (2016) [26] Coecke et al. (2020) [31] Meichanetzidis et al. (2020) [49]
	Disambiguation Question answering	Correia (2021) [51]
	Machine translation	Abbaszade et al (2021) [52]
Quantum-inspired approaches (classical hardware)	Information retrieval	Sordoni et al. (2013) [53] Xie et al. (2015) [54] Li et al. (2018) [55] Jiang et al. (2020) [56]
	Question answering	Zhang et al. (2018) [57] Zhang et al. (2018) [58] Li et al. (2019) [59]
	Sentiment Analysis	Zhang et al. (2018) [60] Zhang et al. (2019) [61] Zhang et al. (2020) [62] Zhang et al. (2021) [62]
Quantum computer approaches (NISQ hardware)	Question answering	Meichanetzidis et al. (2020) [63]
	Sentence classification	Lorenz et al. (2021) [64]
	Machine translation	Vicente (2021) [65]

Pre-QNLP Approaches

Even in its original pre-quantum formulation, the DisCoCat framework has had great success in the literature, as it was used for various linguistic phenomena. For instance, following an experiment proposed by [66], the work of [67] has tested DisCoCat on British National Corpus (BNC) [68] in order to demonstrate the possibility of a practical implementation of such a model opening the way to the production of large scale compositional models. More recently, other aspects connected to DisCoCat model have been investigated, including tasks such as negation and disambiguation. In particular, in [69] the possibility of a model for logical negation has been investigated. The work of [70] instead has proposed a method exploiting the distributional compositional categorical model of meaning to compare and translate sentences in English and Irish using via vector spaces.

Afterwards, the tensor-network based approach on which DisCoCat is based has given rise to numerous derivative approaches. Among these, it is important to mention the study proposed by [71]. It has used the Frobenius algebra [72] to develop a compositional vector-based semantics of subject, and object relative pronouns within a categorical framework

have been supported. The general operations derived from the Frobenius algebra have been used to formalize the operations required to model the semantics of relative pronouns, ranging from relative clauses to modifiers. Again, the feasibility of the approach has been assessed through small-scale experiments on BNC corpus. Another work that has exploited tensor networks following DisCoCat idea has been proposed by [73]. It is a work focused on the disambiguation task. In particular, an algorithm able to improve number of tensor-based compositional distributional models of meaning introducing a step of disambiguation prior to composition has been developed. This approach is based on a methodology already known in the literature for automatic word sense disambiguation [74], which creates unambiguous versions of tensors before these are composed with vectors of nouns to construct vectors for sentences and phrases.

4. Theoretical Quantum Approaches

Since its release, DisCoCat model offers several advantages on a theoretical level; in particular, it has provided a syntax-aware algorithm able to compute the meaning of sentences and phrases using tensor product composition. Although it is originally inspired by some properties and protocols of quantum mechanics [75], its implementation was limited by the high computational cost it required.

Addressing this issue, the work of [26] proposes a way to implement distributional compositional models, such as DisCoCat using quantum computers. The starting assumption is that quantum computation innately fits for managing high dimensional tensor product spaces. Computational problems related to tensor-product-based compositional semantics are addressed by the scalability of quantum systems. In addition, the work proposes an efficient sentence similarity algorithm. These kinds of algorithms are good candidates to have a real implementation on a quantum hardware, since noisy results can be tolerated.

The proposed algorithm exploits the abstract connection between NLP and quantum information modeling density matrices using mixed states of quantum systems. Concerning performance, the quadratic speedup has been demonstrated in sentence similarity task.

However, this performance improvement has only been demonstrated on a theoretical level, since it is based on the assumption that it is possible to use QRAM, which is currently unavailable [21].

An alternative approach born to compensate for this lack has been proposed by [49]. This work theorizes a full-stack pipeline for NLP on a NISQ device exploiting the classical ansatz parameters instead of the never realized QRAM of the original algorithm. Starting from the tensor structure already proposed in the non-NISQ friendly algorithm [26], every step of the theorized NLP pipeline is described in detail, specifying the tools and methods that would be possible to apply to be compliant with a quantum hardware.

In particular, ansätze are used as lexical categories to map DisCoCat models to variational quantum circuits. These categories are connected following the grammar in order to build circuits for arbitrary syntactic units. Two methods have been proposed for optimizing the resulting circuits, aiming at the implementation on NISQ devices. Moreover, the work explores alternative quantum computing models, i.e., continuous- variable, adiabatic and measurement-based.

Although this is an essentially theoretical work that leaves many open issues (for instance, the limitations of CFG to approximate some linguistic phenomena), this method has been followed by an actual implementation on a NISQ device [63]. Even the implementation has been intended as a proof of concept experiment, and if it is tested on a very small scale, it can be considered the first QNLP experiment running on quantum hardware.

The research line proposed by these preliminary works has been continued and expanded in [31]. The vast majority of issues addressed by this study will be subsequently explored in further theoretical and practical work. First, it provides mathematical foundations for QNLP, exploiting diagrammatic theories provided by DisCoCat and CQM. Using these diagrammatic systems of representation, language can be interpreted in terms of

quantum processes. Moreover, the resulting quantum representation of natural language can be embedded in quantum circuit form using the ZX-calculus [76]. The ZX-calculus acts as a translator from linguistic diagrams to quantum circuit. This step is crucial to allow a real implementation on quantum hardware.

This approach relies on the advocated intrinsic quantum-native nature of QNLP, allowing the possibility to merge meaning and structure of language in a single unified quantum model able to represent both syntactic and semantic levels. Another interesting cue triggered by the work of [31] is the potential meaning-awareness of the QNLP, an aspect that puts it in contrast with the vast majority of state-of-the-art machine learning approaches. Different levels of linguistic knowledge would not be more embedded in a “black-box” (hidden layers of a neural network), but they could be rendered explicit through the expressive possibilities of the diagrammatic representation.

Concerning possible applications, the possibility to execute NLP task on NISQ devices is discussed, although no details are presented on whether to deal with an actual implementation. The quantum speedup that the proposed approach should provide is also not demonstrated. However, the possibilities offered by NISQ paradigm are strongly supported. Notice that it is not proposed as a simple alternative to traditional methods, but variational quantum circuits [77] are considered as the best option able to offer a QNLP-friendly way for encoding linguistic data on quantum hardware.

More recently, a novel framework based on tensor contraction to build word representations as quantum states that serve as input to a quantum algorithm has been introduced by [51]. This makes it particularly suitable for classification tasks, such as sentence disambiguation or binary (yes-no) question answering. Starting from this assumption, the work proposes to use an implementation of Grover’s quantum search algorithm to find the answer to a wh-question with quantum speedup using information given directly by the tensor contractions of representations of words, as proposed by the tensor network.

The proposed framework has also been tested on a second NLP task, such as a sentence-meaning disambiguation task. In this case, the advantage of quantum superposition is exploited to store various meanings of ambiguous sentences.

Finally, the diagrammatic language of categorical quantum logic of DisCoCat has given the opportunity to compare the grammatical structures of sentences in different languages.

The study of [52] has been included in this thread. It focuses on developing compositional vector-based semantics of sentences using quantum natural language processing to compare synonymous simple sentences in English and Persian using their parametrized quantum circuits. This approach is based on a quantum long short-term memory model. Even only two languages have been considered, this approach is virtually able to generalize opening up the possibility to include different languages.

5. Quantum-Inspired Approaches

Notice that algorithms described in this section borrow mathematical frameworks from quantum mechanics but they are not designed to run on quantum hardware, but rather on classical computers. In some cases, they could theoretically be used on quantum hardware, which, however, is not currently available. A comparative visualization of the models proposed by the quantum-inspired works described below, including the datasets on which they have been tested; the performance achieved against the reference baselines is shown in the Table 2.

Table 2. Comparison of quantum-inspired approaches running on classical hardware. Since a real comparison is not always possible, only the works that have really compared with benchmark datasets already known in the literature are shown. For each model proposed, the best score obtained for the metric used with respect to the specific dataset is shown in the last column. In brackets, the best score obtained by the baseline with which each approach has been compared is indicated.

Task	Proposed Approach	Dataset	Metric	Score
Information retrieval	Quantum Language Model (QLM) [53]	SJMN	MAP	0.2093 (0.271)
		TREC7-8		0.2254 (0.2243)
		WT10g		0.2264 (0.2146)
		ClueWeb-B		0.1196 (0.1137)
	Quantum Language Model-based Query Expansion (QLM-QE) [55]	TREC 2013 TREC 2014	MAP@10	8.94 (4.91) 14.79 (14.52)
Question answering	Quantum Interference-inspired Neural Matching model (QINM) [56]	ClueWeb-B	MAP	0.134 (0.082)
		Robust-04		0.294 (0.103)
	Neural Network-based Quantum-like Language Model (NNQLM) [57]	TREC-QA	MAP	0.758 (0.678)
		WIKIQA		0.649 (0.510)
		TREC-QA WIKIQA		0.752 (0.678) 0.695 (0.512)
Complex-valued-Network (CNM) [59]	YahooQA	MAP	0.575 (0.395)	
	TREC-QA WIKIQA		0.770 (0.777) 0.674 (0.652)	
Sentiment classification	GQLM [60]	OMD SS-tweet	Accuracy	0.629 (0.614) 0.618 (0.579)
	Quantum-inspired Interactive Networks (QIN) [61]	MELD	Accuracy	0.679 (0.652)
		IEMOCAP		0.376 (0.351)
	TextTN [62]	MR	Accuracy	82.2 (82.3)
		CR		85.7 (85.4)
Subj	95.3 (94.6)			
MPQA	90.4 (90.4)			
BERT + TextTN [62]	SST-2 SST-5	Accuracy	95.3 (96.7) 54.8 (54.7)	

5.1. Information Retrieval

One of the first approaches advocating the use of Quantum Theory for an NLP task has been proposed by [53]. In this work, the quantum probabilistic framework is used to develop a general approach to the information retrieval task. A density matrix has been used to represent texts and term dependencies in a quantum language model (QLM), combining the flexibility of vector spaces with probabilistic calculus. Dependencies are not more represented stochastically as joined probabilities but as superposition events, i.e., additional dimensions.

The experiments have compared the proposed QLM to the baseline unigram language model, based on a bag-of-words approach, in particular Dirichlet smoothing, and to the full dependence version of the non bag-of-words MRF model. Tests have been made on different datasets (SJMN, TREC7-8, WT10g and ClueWeb-B) using the mean average precision (MAP) as metric. QLM consistently achieves results above baseline, with MAP increments varying from 12% to 19% depending on the dataset used.

A similar work [54] has proposed to introduce the quantum entanglement in order to improve the performance of QLM in the information retrieval task. The basic assumption underlying this approach is the equivalence between quantum entanglement (QE) with a traditional retrieval approach, in this case Unconditional Pure Dependence (UDP) [78] used in the Markov Random Field (MRF) retrieval model [79]. Starting from this, the occurrence of QE can be inferred extracting UDP patterns. After that, the QLM has been enhanced using QE, producing a new quantum-entanglement model for information retrieval (QQE). This enhanced model has been shown to achieve better performance than both the traditional model (MRE) and the QML. MAP has been used as metric to evaluate retrieval performance on five collections in the documents' re-ranking task. However, performance has been

evaluated only on a set of custom-created document collections and not on any benchmark dataset, making comparison with other models difficult.

Inspired by these works, another QLM that incorporated a query expansion framework to overcome restrictions due to limited vocabulary has been proposed in [55]. Despite successful applications of QLMs in the field of Information Retrieval, all these approaches are biased by the fact that the training algorithm for the original QLM is not globally convergent. Then, in order to try to reduce these imperfections affecting the final ranking, approaches that use a QLM are based on a minimal vocabulary that typically coincides with query terms. This compromised results in a significant reduction in the performance capabilities possible with QLM. Therefore, incorporating a query expansion framework into QLM opens up the possibility of using a higher number of terms into modeling. The approach is structured as follows: first, an advanced training algorithm is used to generate first-round ranking, then—in order to identify the expansion terms—a single QLM is built for the top returned results. Finally, the final ranking result is generated by applying a QLM with the new training algorithm to the expanded query.

The proposed Quantum Language Model-based Query Expansion (QLM-QE) framework has been tested on TREC datasets showing an improvement when compared to other traditional approaches. In particular, QLM-QE has been compared with four other models to evaluate performance: the Unigram baseline model, RM-HS [80], a query expansion approach inspired by quantum interference, the original QLM proposed by [53], and QMT [81], which is a quantum-like session search model inspired by the Two-State Vector Formalism (TSVF). Performance has been evaluated on TREC datasets, since they are the de facto standard for the performance evaluation for quantum models. nDCG@10 and MAP@10 have been chosen as metrics. QLM-QE has proven to outperform all other models, achieving the result of 10.37%, 15.19% (nDCG@10) and 8.94%, 17.79% (MAP@10) on TREC 2013 and TREC 2014, respectively. The most impressive result has been achieved on TREC 2013, in which there is a 89.81% improvement in MAP@10 score over the baseline and 70.83% compared with the best-performing quantum-model (QMT). However—in spite of these results—further experiments are needed to demonstrate the real benefit of a query expansion approach. In particular, comparisons should be made with classical query expansion approaches and using specific datasets for information retrieval tasks.

More recently, Ref. [56] has proposed another improvement to QLM applied to information retrieval by introducing the interference effects in the neural matching model. The work starts from a revision of neural matching models for retrieval, using a probabilistic quantum property. The proposed model—namely the Quantum Interference-inspired Neural Matching Model (QINM)—applies quantum interference theory to the neural matching model for retrieval. In particular, the proposed methodology construct the probability distribution of a document into the reduced density operator, after which the effective probability distribution is extracted using a N-gram window convolution network. Finally, the query attention mechanism calculates matching features and the final matching score is obtained. A set of experiments have been performed to test the effectiveness of the interference matching information. Performances achieved by QINM on ClueWeb-09 and Robust-04 datasets have shown a significant improvements with respect to other QLMs.

In particular, the model has been compared not only with QLM [53], but also with a neural network-based model (NNQLM), which performs density matrices optimization and learning architecture through two distinct approaches [57], and with a model that uses Quantum Many-body Wave Function and convolutional neural network (QMWF-LM) [58]. QINM has achieved much better results on both datasets using different metrics (MAP, NDCG@20, P@20, ERR@20) showing the advantage in using vectors and neural networks in the retrieval process without focusing only on semantic information but also extracting effective matching features, whose function turns out to be decisive. It is important to note that good performance is highly dependent on taking into account interference effects in relevance judgment process, which is an aspect totally ignored by other QLMs. However, it is essential to consider that the models undergoing this comparison were

not designed specifically for the information retrieval task. Moreover, unlike previous work, in this case the model has been tested not only against other QLMs but also against traditional information retrieval systems, and it has achieved better performance (0.255 using MAP on the Robust-04 dataset compared to 0.418 reached by BM25 [82]). This result is further evidence of the advantage represented by using probability theory to improve retrieval performance.

Summing up, the information retrieval task has largely benefited from quantum-inspired approaches. Given the multidisciplinary nature of the task, using methods and concepts derived from quantum mechanics (in particular quantum probability) has led to the development of several quantum models, exploiting the representation possibilities offered by complex vector spaces (Hilbert space) [53]. This method of representation has been effective for the abstraction and contextualization of information objects such as documents and queries [83]. Different approaches have ranged from using traditional properties, such as query expansion [55] to phenomena such as quantum interference placed in the framework of neural networks [56]. Although QLMs promise the best way to model phenomena affecting information retrieval, such as ambiguity and dynamic changes of context, the research on this area requires more formal rigor both in the design of models with a strong theoretical basis with respect to the task, and in experiments.

5.2. Question Answering

Another task for which the application of quantum approaches has met with particular success is Question Answering (QA). Differently from information retrieval, a QA task is often composed by a question identified by a piece of natural language contained into a sentence or multiple keywords. Candidate answers are usually shorter than documents in the information retrieval task, hence there is low probability of overlapping between question and answer sentences. For these reasons, approaching the QA task using semantic matching via neural network is very common in the literature.

Given this premise, Ref. [57] has proposed to improve the original QLM model designed only for the information retrieval task [53]. A Neural Network-based Quantum-like Language Model (NNQLM) has been developed and applied to the QA task. It utilizes word embedding vectors as the state vectors, from which a density matrix is extracted and integrated into an end-to-end Neural Network (NN) structure. Information about similarities between each question-answer pair is encoded by density matrices in a joint representation. This representation is the source where similarity features are extracted using a Convolutional Neural Network (CNN). The model has been tested on two datasets, TREC-QA, one of the standard benchmark resource for this task [84], and WikiQA [85], an open domain question answering dataset. For the evaluation, the MAP metric has been used, in order to make possible the comparison with previous works for the same task which use same datasets.

Several experiments have been conducted to test NNQLM on both datasets. First, it has been compared to QLM, improving performance by 11.87% on TREC-QA, and by 27.15% on WikiQA using the MAP metric (results are shown in Table 2). These results show how positively the combination of the density matrix and the simple training algorithm can impact performance. A further comparison has been carried out between NNQLM and existing neural network-based approaches. The proposed model has achieved higher scores than most baselines (for instance the MAP score of the strong baseline proposed by [86] by 2.46% on WikiQA). However, NNQLM fails to overcome some of the baselines [87,88]. In fact—even if the model is able to learn similarity patterns—it does not take into account aspects such as structures and relations of sentences and this affects the overall performance.

This work is continued and extended in [58] introducing Quantum Many-body Wave Function (QMWF) for language modeling (LM) to address QA task. QMWF-LM improves the representation space of QLMs being able to represent complex word interactions, using semantic basis vector corresponding to multiple word meanings. In addition, this work aims to solve the problem of the connection between QLMs and CNN. In particular,

QMWF-LM uses a series of derivations based on projection and tensor decomposition to demonstrate that quantum representation and matching can be implemented by the CNN with a product pooling. Based on this assumption, QMWF-LM is presented as an efficient algorithm to represent and match the text/sentence pairs, and is therefore well suitable for QA tasks. QMWF-LM has been tested on different QA datasets; it has proven to be able to outperform quantum LM counterparts (i.e., QLM and NNQLM) and even to achieve better scores, whether compared with traditional CNN-based approaches on same datasets. Datasets taken into account are—similarly to previous works—TREC and WikiQA. Moreover, a third dataset has been added, YahooQA, typically used as benchmark dataset for community-based QA.

Although QMWF-LM significantly improves the performance of the original QLM (0.695% MAP with respect to 0.512% on WikiQA) some clarifications should be made. The low performance of QLM is easily explained since—unlike later approaches (NNQLM and QMWF-LM)—it is trained in an unsupervised manner. Whereas, the better performance obtained by QMWF-LM than by NNQLM is motivated by the fact that the latter does not take into account word interactions, using embedding vectors as input and a CNN to train the density matrix. Finally, QMWF-LM has been also compared with other classical CNN-based QA models [86,88], showing a comparable or even better performance.

Another work addressing QA proposes an even different approach based on a complex-valued framework [59]. Starting from a neural network built using density matrices proposed in [57], a novel quantum-theoretic framework to model language in a way closer to a cognitive point of view has been proposed. In the proposed framework, different linguistic units are modeled as quantum states using quantum probability, i.e., the mathematical framework of quantum physics aiming at model uncertainly on a unified Hilbert space with well-defined mathematical constraints and explicit physical meaning. These linguistic units are represented as complex-values vectors to preserve physical properties. The length represents the relative weight of the word and the direction is viewed as a superposition state. The superposition state is further represented in an amplitude-phase manner, in which amplitude is consistent with the lexical aspects, while phases represent semantic ones. The framework has been implemented using a complex-valued-network (CNM) and applied to QA task to be evaluated. Unlike previous approaches, in this case no CNN or RNN architectures are used, because of the difficulty to interpret structures of convolutional kernels and recurrent cells. The proposed CNM offers a simpler alternative in terms of understanding.

CNM has been tested on benchmark QA datasets (TREC and QikiQA) showing comparable performance to baselines. In particular, it achieves better performance of most traditional CNN and LSTM-based models, characterized by a more complicated structure and a larger number of parameters. In addition, it outperforms other QLMs on both datasets (in particular, it improves NNQLM MAP score by 3.88%), supporting the validity of the quantum theoretical framework.

Quantum approaches developed so far for question answering are varied and they have addressed the task in very different ways. In [57], for the first time, the original QLM model has been extended using neural network architectures to be applied to question answering. New density matrices, for single sentences and sentence pairs, based on word embeddings have been integrated into the architecture for an effective joint training. However, this approach does not allow one to easily represent complex interactions between words, since it considers a compound word as a direct addition of the representation vectors or subspaces of the single words involved. Moreover, the desired union between quantum models and neural network remains hinted at but not formally approached. To deepen the intrinsic motivations to bridge neural network and quantum models, in [58], the analogy between the quantum many-body system and the language modeling has been proposed, trying to merge a quantum many-body wave function-inspired model with a neural network. A further investigation of the possibility offered by a robust quantum-

inspired neural architecture in a higher-dimension Hilbert space has been conducted by [59].

5.3. Sentiment Analysis

Another NLP task that has benefited from many quantum-based approaches is text classification, and in particular, sentiment analysis. A method that uses features derived from quantum probability theory to perform sentiment classification has been developed by [60]. It is an unsupervised sentiment analysis approach based on a density matrix. The approach is structured as follows: two sentiment dictionaries are ad hoc created, using extended QLM density matrices of documents, and dictionaries are generated; the sentiment is the result of the similarity between dictionaries and documents calculated using quantum relative entropy. The viability of this approach to sentiment analysis has been evaluated using two datasets well-known in the literature, the Obama-McCain Debate (OMD) dataset and Sentiment Strength Twitter dataset (SS-Tweet). The results have demonstrated higher performance of this approach compared to various baselines.

This initial work has evolved [61], introducing a new approach using quantum-inspired interactive networks. Aiming at better modeling interactions and dependency dynamics between words in existing sentiment analysis approaches, authors propose a quantum-inspired interactive networks (QIN), combining quantum theory with the long short-term memory (LSTM) neural network. Correlations between words captured by the density matrix are used as an input of long LSTM. In detail, this method can be able to learn such interaction dynamics, using a density matrix-based convolutional neural network (DM-CNN) to capture intra-utterance correlations between words. A strong-weak model based on quantum measurement theory is also used to extract inter-utterance correlations and measure the influence between speakers through the utterance. The influence is integrated in an LSTM which uses textual features as inputs. Starting from the hidden states of the LSTM, affective states for each utterance are determined using a softmax function. QIN has been compared to several baselines ranging from CNN [89] to attention-based LSTM and Contextual/Hierarchical biLSTM [90]. Experiments have been carried out on two datasets with different types of annotations: MELD (3-class sentiments; 7-class emotions) and IEMOCAP (9-class emotions). In sentiment classification (3-class MELD dataset), QIN achieves higher performance than all classical models, since they do not consider contextual dependencies among utterances ignoring that the whole meaning of a utterance is influenced by preceding ones. Increasing the complexity of the task and introducing more classes (MELD 7-class and IEMOCAP 9-class), QIN continues to get the best results, but the gap narrows, settling on a 7.1% accuracy improvement over biLSTM. BiLSTM successfully achieves superior performance to other baselines, since it is able to extract contextual features, taking utterances as inputs.

More recently, a further improvement of the original work has been proposed in [62] introducing a new architecture based on a tensor network to improve performance, interpretability and expressive power. In this work, a model based on tensor network (TN) able to obtain state-of-the-art results in sentiment classification task has been proposed. The model, called TextTN, exploits the great expressive power of TNs [91] to build a tensor network-based probabilistic model for natural language representation and classification. In order to learn and classify texts, TextTN encodes each high-dimensional word vector in a probabilistic space by a generative TN (word-GTN). Then, a discriminative TN (sentence-DTN) is used to classify sentences. It trains a TN for each sentence using as input word vectors extracted by word-GTNs. Sentence-DTN is also devoted to handle word interaction by the joint effect of different words for the later class predication by the loss functions. Concerning the learning algorithm, an all-function training process in the sentence-DTN has been proposed to improve the stability of TextTN. TextTN has been extensively evaluated using main sentiment text classification datasets (MR, Subj, CR and MPQA). In the comparison with classic benchmarks models, it has achieved better results than CNN [89] on all four datasets, with accuracy improvements ranging from a minimum

of 0.7% to 1.9%. The comparison with the state-of-the-art self attention model proposed by [92] has returned a 0.3% and 0.7% increase, respectively, on CR and Subj datasets, while the score has remained identical on MPQA. Moreover, experiments have been carried out using TextTN in combination with pre-training word vectors from BERT on SST-2 [93] and SST-5 [94] tasks. Concerning SST-2, BERT+TextTN performs better than BERT (accuracy improved by 0.7% for SST-2 and by 1.9% for SST-5) and is comparable with ELECTRA [4].

In summary, quantum-inspired models for sentiment analysis can be considered a generalization of classical approaches. Ref. [60] proposes first an improved version of QLM for twitter sentiment analysis using the quantum probability theory. The work is extended and expanded with comparative studies and the use of different datasets, annotated with both polarity and emotion in [58]. This work combines a density matrix-based CNN with a model inspired by quantum measurement theory. An even more sophisticated quantum model, which makes even better use of the expressive potential of tensor networks, is the one proposed in [62]. Notice that, although these models achieve impressive results, there are some open issues. It is quite difficult to make proper comparison between different quantum-inspired approaches for sentiment analysis. First, experiments are carried out only in comparison with traditional models. Moreover, every approach is evaluated on different datasets, then the datasets can have different types of annotations with varying numbers of classes. This is inherent in the articulated nature of the task, but it makes the performance strongly dataset-dependent.

It is worth mentioning other works focused on tasks that have received less attention in the literature. In [95], the possibility to boost neural language models based on deep-learning using QLM has been suggested. This approach trains a quantum-enhanced Long LSTM to perform a PoS tagging task using numerical simulations. Among the advantages of the implemented quantum LSTM is the reduced use of parameters, which is less than half of the traditional one, but keeps the same performance. In addition, this work takes the first steps toward the creation of a quantum-enhanced Transformer model, used to perform sentiment analysis task. A preliminary experiment has been carried out on a IMDB dataset, although no comparative results are reported. Notice that the work does not propose the implementation of a proper quantum transformer, in which quantum circuit acts on qubits similarly to self attention and positional embedding. Rather, a quantum alternative for Transformers' arithmetic operations is offered, even because a full implementation exceeds the current computational possibilities of the available quantum hardware.

Finally, the work proposed by [96] has explored the possibility to apply a QLM to a speech recognition task. Although this work is intended as a "proof of concept" and it does not have a full implementation, an evaluation phase has been carried out on the TIMIT dataset. The proposed QLM has been compared with two N-gram implementations and two RNN models, achieving performance comparable with state-of-the-art.

6. Quantum Computer Approaches

The following section presents works using real quantum hardware. All work listed below use NISQ computers [33]. In Table 3, some specifications of hardware and experiments are provided.

The first implementation of an NLP task on NISQ hardware has been proposed by [63], following theoretical methods proposed in [49]. Conceptual and mathematical foundations on which these work are based are described in [31]. It uses DisCoCat, adopting the paradigm of Parameterized quantum circuits as machine learning models [77]. The core of the proposed approach is considering DisCoCat as a tensor network model of natural language meaning [97]. Meanings of words are encoded as co-occurrence frequencies or other word-embeddings produced by neural network. These tensor networks can be represented as string diagrams [98], composed of boxes with input and output wires, each of which carries a type. Boxes are composed to form process networks by wiring outputs to inputs by ensuring types are respected. Output-only processes are called states and input-only processes are called effects.

The approach has been tested on a question answering task using a labeled dataset composed by 30 randomly generated sentences using a finite vocabulary and a context-free grammar (CFG) [99] and represented as a syntax tree. These sentences are converted into DisCoCat diagrams. An optimization (or training) step is needed in order to match the predicted labels with labels in the training set. The optimization has been performed using SPSA (Simultaneous Perturbation Stochastic Approximation) [100], which has been proven to perform well in noisy settings. This approach reformulates the problem as a supervised learning task of binary classification for sentences, i.e., a special case of question answering already known in the literature [101,102]. Before each circuit can be run on a backend, it has been compiled using the quantum compiler $t|ket\rangle^{TM}$ [103]. The quantum compilation operates in this way: given a circuit and a device, quantum operations are decomposed in terms of devices native gateset, then the quantum circuit is made compatible with the device topology. In addition, the compiler tries to minimize the most noisy operations. After that, the circuits run on two IBMQ quantum computers `ibmq_montreal` and `ibmq_toronto`, whose quantum volume is 32 for both, with a maximum allowed number of shots equal to 2^{13} .

Concerning results, first a classical simulation has been performed. For the QA task, sentence circuits have been evaluated using classical hardware. For this simulation, the entire corpus of 30 sentences with a vocabulary consisting of 7 words has been taken into account. The experiments running on NISQ device has instead exploited only 16 sentences with a vocabulary of 6 words. The best results achieved reach a train error of 12.5% and a test error of 37.5% on `ibmq_toronto`; values decrease further on `ibmq_montreal`, where the train error is null, while the test error stops at 37.5%.

Table 3. Comparison of quantum experiments running on NISQ hardware. Note that not all details are given in the description of the experiments performed. Concerning NISQ devices, all machines used have 5 qubits available and a quantum volume equal to 32.

Task	Device	Dataset	Train Error	Test Error
Question answering	simulation	30 sentences 7 words vocabulary	/	/
	<code>ibmq_toronto</code>	16 sentences	12.5%	37.5%
	<code>ibmq_montreal</code>	6 words vocabulary	0	37.5%
Meaning classification (MC)	simulation	130 sentences	16.9%	20.2%
	<code>ibmq_bogota</code>	17 words vocabulary	/	16.7%
RelPron (RP)	simulation	105 sentences	9.4%	27.7%
	<code>ibmq_bogota</code>	115 words vocabulary	/	32.3%

In [64], the first full experiments at a medium-scale, focused on two NLP tasks running on quantum hardware, have been performed. This work starts from earlier work [63] and—differently from previous approaches—does not intend to provide evidence of the so-called “quantum advantage” in terms of performance. Therefore, these experiments aim to explore the challenges and limitations of training and running an NLP model on a NISQ device. In addition, this work contributes to shifting traditional approaches to NLP tasks to a quantum-friendly form, proposing a quantum pipeline equivalent to the classical one. Two tasks are proposed, and both of them are structured as binary classification problems. The first one, namely the classification task (MC), uses a dataset of 130 sentences plain-syntax generated from a fixed vocabulary using a simple CFG that can refer to one of two possible topics, food or IT. For the second task, 105 noun phrases are extracted from the RelPron dataset [104], and the goal of the model is to predict whether a noun phrase contains a subject-based or an object-based relative clause. This task, identified using the acronym RP, is more challenging since it requires some syntactic awareness from the model, so it is a very reasonable choice for testing DisCoCat. Notice that, even these datasets can appear small or simple, as they already reach the limits of the currently available quantum hardware. Using DisCoCat, every different grammatical structure of a sentence

is mapped to different quantum circuits. A simplified version of the quantum pipeline used in this work is shown in Figure 3. In detail, in order to propose a full NLP pipeline equivalent to classical ones, sentences need to be represented in a tree form consistent using a parser. A parser, based on categorial grammar [105], has been used to obtain representation consistent with DisCoCat. Subsequently, every diagram is mapped to a specific quantum circuit. The mapping is defined by a conjunction of choices formulated as an ansatz, allowing to make choices based on the specific hardware. Every NISQ machine has a different sets of native gates indeed, of which some are more prone to error than others. These steps have been implemented using the python implementation of DisCoCAT [106]. To estimate the expectations of a hypothetical noise-free quantum device, a simulation has been performed. According to this simulation, the train and test errors reported after 500 iterations for the MC task are 16.9% and 20.2%, respectively. For the second task on the RELPRON dataset, RP train errors are equal to 9.4%, while those of the test reach 27.7%. Higher test error value for the RP task—at equal task complexity—are motivated by a broader vocabulary (115 words) with respect to the MC one (17 words). Once each sentence has been represented by a quantum circuit according the the chosen ansatz, circuits are converted into machine specific instructions using $t|tet\rangle^{TM}$ compiler. After that, the quantum device `ibmq_bogota` executes the circuit n_{shots} times. This IBM machine is a superconducting quantum computing device with 5 qubits and a quantum volume of 32. In particular, initial states are prepared for each run, gates are applied and all qubits are measured. This step returns the count of the shots for all qubits. Every time the error cost is calculated, compiled circuits corresponding to all sentences are sent to the device as a single job. After that, each circuit is run the the maximum possible number of time offered by the machine (2^{13} in this case). For every sentence, estimations of relative frequencies are calculated in the order of a post-processing step aimed at calculating task-specific results. In order to minimize the impact of quantum noise, the experiments has been performed on a single run (100 iterations for the MC task with a exclusive access of `ibmq_bogota` with 12 h of runtime and 130 iterations for the RP task in a sharing mode with 72 h or runtime). Results demonstrate a test error of 16.7% for the MC task and 32.3% for the RP one, with F-score of 0.85 and 0.75. Although it is impossible to make an objective comparison with the simulator, these can be considered good results concerning the size of the datasets, factors affecting NISQ devices and different conditions in which experiments have been carried out. The simulation has been run on classical hardware without noise during many averaged runs, while quantum computation has been performed on a single run on a device affected by noise.

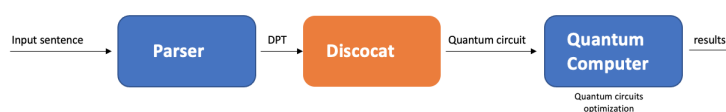


Figure 3. A simplified version of the pipeline proposed by [64].

In [65], a preliminary experiment focused on machine translation using DisCoCat has been proposed. The goal of the experiment is the possibility of a quantum-like approach for language understanding in different languages.

This is the first work trying to use DisCoCat for a language other than English. In particular, simple, complex and negative sentences in two languages (Spanish and English) have been converted into DisCoCat diagrams, then into quantum circuits. Negative sentences do not succeed to be managed from the system even if DisCoCat model is capable of describing negative sentences and constructing diagrams. However, although some solutions have been proposed in the literature, the possibility of mapping negations using functors is still an open issue [107].

After that, these sentences have been encoded in an IBMQ quantum machine using python language. In particular, using an optimized model (IBM Q NISQ backend) has

demonstrated that it can answer factoid questions related to an input vocabulary. According to the theoretical approaches proposed in the literature [63], the score achieved by IBMQ is close to 1, indicating the correctness of the statement. The evaluation phase has shown that the system is able to give consistent answers handling simple sentences. The results are also quite good, as the complexity and length of the sentences increase. Evaluation phase show a sentence similarity between the two languages, reaching a value of 95%. Notice that, although the work claims to have used a NISQ device, no specific information is provided about the implementation, machine type or dataset in detail.

7. Discussion

As can be observed from the Table 1, there has been an evolution in the tasks affected by the QNLP. The possibilities offered by the compositional distributional model has opened up a line of research that initially (in a pre-quantum era) has focused on very specific tasks and critical aspects of NLP. On the one hand, the initial interests of scholars have addressed the issue of a formal syntax-semantic representation, testing the DisCoCat framework of well-known corpora such as the BNC. These interests in formal aspects of language have ranged from modeling structures such as logical negation to disambiguation. Many of these aspects have already been dealt with in the linguistic field with Lambek's pregroup grammar and within the broader framework of categorical grammars [108,109].

These subtle linguistic issues have remained the main topic moving to the early theoretical quantum approaches. In these works, first, QNLP algorithms have been proposed, theoretically demonstrating the advantages at the methodological level and in terms of performance. The possibility to successfully execute DisCoCat on a quantum hardware benefiting from quantum speedup has first been addressed in [26].

In order to overcome the big shortcoming of this first attempt, i.e., a total dependence on the theorized but never realized QRAM, alternative solutions have been proposed. In [31], a family of quantum machine learning algorithms—namely variational quantum circuits—have been used to indirectly encode data on the quantum computer. A different approach is the one proposed in [49]; to avoid the need of QRAM entirely, the classical ansatz parameters are exploited to encode distributional embedding. The feasibility of the latter approach has been confirmed, since the proposed pipeline has actually been implemented on a NISQ device—albeit with limitations and on a small scale—and it has achieved good performance in the question answering task.

Concerning hybrid quantum-inspired approaches running on classical hardware, this is the sub-field that has attracted the most interest in the literature so far. On the one hand, the reason is the physical possibility of being able to perform experiments without having to rely on quantum machines. On the other hand, these are the only approaches that have had the possibility of being compared with benchmark datasets in the literature and for which performance can be estimated according to classical metrics that are well known in NLP tasks.

These hybrid approaches have arisen from the need to study the machine learning and artificial intelligence problems using a point of view borrowed from quantum theory.

The first QLM proposed has taken into account the quantum probability theory to overcome the limitations of traditional LMs, based on classic probability theory, whose number of needed parameters grows progressively with the complexity of word dependencies. The natural field in which to apply this QLM has been information retrieval. The original QLM is based on encoding the probability measurement for both single words and compound words using a density matrix with a fixed dimensionality. Derivative works have extended the approach by introducing features such as query expansion or quantum interference. The success of this approach applied to ad hoc information retrieval task, which is capable of achieving performance comparable to classical LMs, has made it the benchmark model used to evaluate the performance of all subsequent quantum-inspired works. There are many reasons to apply the first quantum-inspired to the Information Retrieval task. Firstly, there is a historical reason. Information Retrieval has been the first

task for which the possibility to apply the mathematical framework of quantum mechanics in Hilbert space has been theorized [83]. In addition, there is an immediate motivation on a practical level, since the quantum advantage impacts mainly on tasks whose cost tends to be exponential, such as Information Retrieval. After that, several studies in the literature have demonstrated that quantum speedup can more likely affect NLP tasks related to similarity calculations. This has meant that subsequent work has focused not only on the similarity of the query and the documents (Information Retrieval) but also on the similarity of the question and the answers (question answering) and the similarity of the sentences (text classification).

Concerning Question Answering, models that combine QLM and end-to-end neural networks to exploit the density matrix better have been developed (NNLQM). Subsequently, other work has highlighted the inability of early models to apply the fundamental theoretical connections between quantum theory and neural networks in the language modeling process. Therefore, more sophisticated question answering models have been proposed in order to model the complex interactions among ambiguous words with multiple meanings using properties such as the Quantum Many-body Wave Function. Moving to the text classification task, In this case, the QLMs have focused on a specific case, i.e., sentiment analysis task. The interest, as well as the difficulty of this task, lies in the interactions between the utterances and the variability of the datasets. Sentiment datasets can have different annotations based on the emotions/sentiments involved in the analysis. The task has been approached in different ways, using a density matrix-based convolutional network to identify interactions within each utterance or formalizing a tensor network method for natural language representation.

Notice that almost all of these models are based on a density matrix defined in a quantum probabilistic space. Density matrix has proven to be an effective way in representing and modeling language in different NLP tasks, encoding more semantic dependencies with respect to classic word embeddings. Moreover, NLP tasks—although well known in the literature—are relatively simple and structured in a comparable manner. Concerning evaluation, although some models claim to outperform deep learning-based approaches, a proper comparative analysis between QLMs is currently not possible. In fact, there is still no agreement on either the baseline dataset or the metrics chosen by QLMs; each model has been tested against different baselines and compared with a different set of other models.

The experimental stage of QNLP is even more evident, moving to approaches running on real quantum hardware. Significantly few approaches have been developed so far, and there is a further simplification concerning NLP tasks. The tasks remain roughly the same (question answering [63], text classification [64] and a preliminary experiment on translation of simple sentences [65]; the experiments have been carried out on custom small-scale datasets with simple sentences and a minimal vocabulary. Even the largest dataset used so far in a NISQ experiment (100 sentences for a classification task in [64]) remains on dimension enormously smaller if compared to that of deep learning-based architectures. Notice that at the time of the listed work, NISQ devices can have 50-100 qubits. Currently, the most powerful IBM Quantum machines reach 127 qubits (source: <https://www.ibm.com/quantum-computing/systems/> accessed on 1 June 2022).

Although these limitations make difficult to compare quantum-hardware approaches with classical deep-learning ones, as hardware power grows, this drawback may be removed. It is highly feasible that, in the future, NISQ devices will be able to work on the benchmark datasets well known in NLP community.

Notice that hardware limits of QNLP are the same shared with the whole field of quantum computing (i.e., unrealized QRAM, the limited number of qubits, lack of fault-tolerant quantum machine). For instance, the lack of the theorized QRAM has forced scholars to find alternative approaches, such as ansatz circuits, which have been proven to achieve good results for the specific task in the parameterization of word meanings [64]. However, this approach forces a preliminary step of choosing an optimal circuit to be performed each time for each specific task. The possibility of generalizing using ansatzes is

far from being proven yet. This makes this type of approach a difficult path to follow in the view of general performance of a model not limited to a single task. The small number of qubits available instead constrains a very low limit on the number of sentences, their length and the overall span of vocabulary. A higher number of sentence or a larger vocabulary imply a scalability issue, since it means a higher dimensional parameter space. This aspect is somewhat inconsistent with the supposed “quantum native” status of QNLP; however, it is only partially caused by the limitations of current quantum hardware architectures.

Indeed, the reasons why approaches running on quantum hardware cannot currently offer a viable alternative to classical models lie not only in technical reasons. For example, solving the above scalability problem is not enough to rely on higher-performance hardware. The process that allows converting natural language sentences into their diagrammatic representation and then into real quantum circuits theorized in [49] and implemented in [63] works only for particular cases because the system’s internal grammar is capable of correctly handling the representation of only specific sentences. Moreover, there are several open issues concerning the implementation of logical operators used in diagrammatic language to deal with specific language structures.

As a summarization of this overview, this field of research has appeared to be at a very preliminary stage. Many conceptual and mathematical foundations have been pointed out in the first works that appeared in literature; many of these aspects have found application in quantum-inspired solutions evaluated on classical hardware, revealing, albeit in an embryonic way, the enormous potentialities of the QNLP. Only very recently, a reduced subset of these quantum-inspired solutions has found a way to be tested on NISQ quantum hardware, showing promising results although operating on small datasets and simplified scenarios, and not considering at all how to exhaustively assess a quantum speed-up for QNLP. To point out some elements worthy of future investigations, first of all, from an experimental point of view, relevant aspects affecting the performances of QNLP approaches regard how to (i) benchmark the optimisation algorithms, (ii) choose ansatz solutions, training methods and various hyper-parameters, (iii) assess the possible relationship between corpus size, wire dimensionality and generalization. This will also open up an exploratory arena for future work on assessing trade-offs of performance achieved with different ansatz families and optimization parameters in a specific task versus general performance on many tasks. On the other hand, from a theoretical perspective, how much the claims held out by the QNLP are valid is yet to be proven at a purely linguistic level. Although these quantum models promise to reduce the bottleneck of the need for extensive annotated resources, it is also true that they require a grammar to describe the structures of the language under analysis. Such grammar must be formalized in a formal logical language and is based on the CFGs mentioned above, whose expressive potential has never been tested on such a large scale. In addition, the low uptake of CFGs in NLP means that there are a small number of tools, such as parsers or pos-taggers, which are currently valid only for specific cases and for small data. This raises questions about the scalability of the approach and practical portability, risking creating a future scenario in which resources (i.e., grammars that formalize the language) are available only to rich-resource languages.

8. Conclusions

Although QNLP is still a new area of research and applications have focused on relatively simple tasks so far, it has already shown numerous opportunities and advantages in many ways.

First of all, from a theoretical point of view, the mechanics of natural language could be better handled by a quantum-based approach: this is the so-called “quantum native” view of QNLP, supported by many studies. Therefore, quantum language models would be better suited to understand and describe natural language phenomena in a way that is more aligned with real human cognitive processes. However, although this claim is fascinating, there is currently no concrete evidence to support it except in managing specific cases, i.e., simple sentences with a limited and controlled vocabulary. This reflects neither

the actual way of learning human language nor its production. From an applicative point of view, probably a solution could come in the future at the increased power and storage capacity of quantum computers that could be the most realistic and suitable option to deal with increasingly large vector spaces. From a theoretical side, introducing different types of sentences with different construction using CFG and Prgroups is costly, since, in some cases, it means that these resources have to be built from scratch. In addition, there are also some concerns about the ability of these kinds of grammars to be able to approximate all kinds of linguistic phenomena

Concerning the so-called “quantum advantage”, similar or even better performance has been achieved by some QNLP models running on classical hardware in various tasks, compared to state-of-the-art baselines. As an example of the application of these statements, QNLP models have been applied to manage aspects of NLP, which have always been critical to deal with the classical probabilistic models, such as interference phenomenon in information retrieval, term dependencies or ambiguity resolution.

However, the performance boost offered by the quantum speedup have been only theoretically demonstrated, relying on a quantum hardware that can take the benefit of still unrealized and expensive QRAM. Although alternative solutions have been found to make up for this shortcoming, implementations on NLP tasks cannot yet be remotely comparable to the classic ones well known in the literature. The data on which these approaches have been tested are in fact very small (about 100 natural language sentences in medium-scale experiments) due to current limitations imposed by NISQ computers and for very restricted tasks due to the need of an ad hoc created grammar for the sentence involved in the experiment.

Concerning future work, the desirable end of dependence on large datasets and complex models with a huge number of parameters can undoubtedly be addressed by the QNLP, but many issues remain open. From a theoretical point of view, there is still the question of scalability in building CFGs underlying the models and the possibility of creating compelling formal descriptions for different languages. Instead, from an application point of view, it could regard the usage of more significant real-world data as well as the implementation of more complex QNLP tasks such as sentence similarity and all those tasks involving ambiguities that may benefit most from the quantum native approach of QNLP. However, although it will presumably take years before QNLP approaches can be executed on a large scale, as is the case today for deep learning-based models, QNLP can offer the unique opportunity to use quantum properties to deal with challenging language phenomena in a manner more similar to how natural language works. This has been demonstrated by the successes of using quantum superposition to model uncertainties and ambiguity in language or entanglement to describe both composition and distribution of syntax and semantics effectively.

Author Contributions: Conceptualization, R.G. and M.E.; data curation, R.G.; formal analysis, R.G.; funding acquisition, G.D.P.; investigation, R.G.; methodology, R.G. and M.E.; project administration, M.E.; supervision, G.D.P. and M.E.; writing—original draft preparation, R.G.; writing—review and editing, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
2. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding, In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Minneapolis, MN, 2019; pp. 4171–4186. [[CrossRef](#)]
3. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
4. Clark, K.; Luong, M.T.; Le, Q.V.; Manning, C.D. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In Proceedings of the ICLR, Addis Ababa, Ethiopia, 26–30 April 2020.
5. Zhu, J.; Xia, Y.; Wu, L.; He, D.; Qin, T.; Zhou, W.; Li, H.; Liu, T. Incorporating BERT into Neural Machine Translation. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
6. Sun, C.; Qiu, X.; Xu, Y.; Huang, X. How to fine-tune bert for text classification? In Proceedings of the China National Conference on Chinese Computational Linguistics, Kunming, China, 18–20 October 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 194–206.
7. Lee, K.; He, L.; Lewis, M.; Zettlemoyer, L. End-to-end Neural Coreference Resolution. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; Association for Computational Linguistics: Copenhagen, Denmark, 2017; pp. 188–197. [[CrossRef](#)]
8. Guarasci, R.; Minutolo, A.; Damiano, E.; De Pietro, G.; Fujita, H.; Esposito, M. ELECTRA for Neural Coreference Resolution in Italian. *IEEE Access* **2021**, *9*, 115643–115654. [[CrossRef](#)]
9. Guarasci, R.; Silvestri, S.; De Pietro, G.; Fujita, H.; Esposito, M. BERT syntactic transfer: A computational experiment on Italian, French and English languages. *Comput. Speech Lang.* **2022**, *71*, 101261. [[CrossRef](#)]
10. Guarasci, R.; Silvestri, S.; De Pietro, G.; Fujita, H.; Esposito, M. Assessing BERT’s ability to learn Italian syntax: A study on null-subject and agreement phenomena. *J. Ambient. Intell. Humaniz. Comput. Volume 12* **2021**, *30*, 1–15. [[CrossRef](#)]
11. Chi, E.A.; Hewitt, J.; Manning, C.D. Finding universal grammatical relations in multilingual BERT. *arXiv* **2020**, arXiv:2005.04511.
12. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds. Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 1877–1901.
13. Floridi, L.; Chiriatti, M. GPT-3: Its nature, scope, limits, and consequences. *Minds Mach.* **2020**, *30*, 681–694. [[CrossRef](#)]
14. Jawahar, G.; Sagot, B.; Seddah, D. What does BERT learn about the structure of language? In Proceedings of the ACL 2019-57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019.
15. Liu, X.; Zheng, Y.; Du, Z.; Ding, M.; Qian, Y.; Yang, Z.; Tang, J. GPT understands, too. *arXiv* **2021**, arXiv:2103.10385.
16. Rogers, A.; Kovaleva, O.; Rumshisky, A. A primer in bertology: What we know about how bert works. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 842–866. [[CrossRef](#)]
17. Jiang, Z.; Xu, F.F.; Araki, J.; Neubig, G. How can we know what language models know? *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 423–438. [[CrossRef](#)]
18. Nielsen, M.A.; Chuang, I. Quantum computation and quantum information. *Am. J. Phys.* **2002**, *70*, 558. [[CrossRef](#)]
19. Kazem, B.R.; Saleh, M.B. The Effect of Pauli gates on the superposition for four-qubit in Bloch sphere. *J. Kerbala Univ.* **2020**, *18*, 33–46.
20. Ben-David, S.; Childs, A.M.; Gilyén, A.; Kretschmer, W.; Podder, S.; Wang, D. Symmetries, graph properties, and quantum speedups. In Proceedings of the 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), Durham, NC, USA, 16–19 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 649–660.
21. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195–202. [[CrossRef](#)] [[PubMed](#)]
22. Khan, T.M.; Robles-Kelly, A. Machine learning: Quantum vs classical. *IEEE Access* **2020**, *8*, 219275–219294. [[CrossRef](#)]
23. Dunjko, V.; Wittek, P. A non-review of Quantum Machine Learning: Trends and explorations. *Quantum Views* **2020**, *4*, 32. [[CrossRef](#)]
24. Coecke, B.; de Felice, G.; Meichanetzidis, K.; Toumi, A.; Gogioso, S.; Chiappori, N. Quantum Natural Language Processing. 2020. <http://www.cs.ox.ac.uk/people/bob.coecke/QNLP-ACT.pdf> (accessed on 1 June 2022)
25. Coecke, B.; Sadrzadeh, M.; Clark, S. Mathematical foundations for a compositional distributional model of meaning. *arXiv* **2010**, arXiv:1003.4394.
26. Zeng, W.; Coecke, B. Quantum Algorithms for Compositional Natural Language Processing. *Electron. Proc. Theor. Comput. Sci.* **2016**, *221*, 67–75. [[CrossRef](#)]
27. Ritter, M.B. Near-term quantum algorithms for quantum many-body systems. *Proc. J. Phys. Conf. Ser. Iop Publ.* **2019**, *1290*, 012003. [[CrossRef](#)]

28. Wu, S.; Li, J.; Zhang, P.; Zhang, Y. Natural Language Processing Meets Quantum Physics: A Survey and Categorization. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online, 7–11 November 2021; pp. 3172–3182.
29. Dunbar, G. Towards a cognitive analysis of polysemy, ambiguity, and vagueness. *Gruyter* **2001**, *12*, 1–14. [[CrossRef](#)]
30. Gries, S.T. Polysemy. *Cognitive Linguistics—Key Topics*; De Gruyter Mouton: Berlin, Germany, 2019; pp. 23–43.
31. Coecke, B.; de Felice, G.; Meichanetzidis, K.; Toumi, A. Foundations for Near-Term Quantum Natural Language Processing. *arXiv* **2020**, arXiv:2012.03755.
32. von Neumann, J. Mathematical foundations of quantum mechanics. In *Investigations in Physics*; Princeton University Press: Princeton, NJ, USA, 1955.
33. Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79. [[CrossRef](#)]
34. Giovannetti, V.; Lloyd, S.; Maccone, L. Quantum random access memory. *Phys. Rev. Lett.* **2008**, *100*, 160501. [[CrossRef](#)] [[PubMed](#)]
35. Grover, L.K. A fast quantum mechanical algorithm for database search. In Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, Philadelphia, PA, USA, 22–24 May 1996; pp. 212–219.
36. Coecke, B. Kindergarten quantum mechanics: Lecture notes. *Aip Conf. Proc.* **2006**, *810*, 81–98.
37. Coecke, B.; Kissinger, A. Picturing quantum processes. In Proceedings of the International Conference on Theory and Application of Diagrams, Edinburgh, UK, 18–22 June 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 28–31.
38. Harris, Z.S. Distributional structure. *Word* **1954**, *10*, 146–162. [[CrossRef](#)]
39. Steedman, M.; Baldridge, J. Combinatory categorial grammar. In *Non-Transformational Syntax: Formal and Explicit Models of Grammar*; Wiley-Blackwell: Hoboken, NJ, USA, 2011; pp. 181–224.
40. Buhrmester, V.; Münch, D.; Arens, M. Analysis of explainers of black box deep neural networks for computer vision: A survey. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 966–989. [[CrossRef](#)]
41. Montague, R. Universal grammar. In *Formal Philosophy*; Yale University Press: London, UK, 1974, Volume 1970, pp. 222–246.
42. Ajdukiewicz, K. Die syntaktische Konnexität. *Stud. Philos.* **1935**, *1*, 1–27.
43. Bar-Hillel, Y. A quasi-arithmetical notation for syntactic description. *Language* **1953**, *29*, 47–58. [[CrossRef](#)]
44. Lambek, J. The mathematics of sentence structure. *Am. Math. Mon.* **1958**, *65*, 154–170. [[CrossRef](#)]
45. Lambek, J. Type grammar revisited. In Proceedings of the International Conference on Logical Aspects of Computational Linguistics, Nancy, France, 22–24 September 1997; Springer: Berlin/Heidelberg, Germany, 1997; pp. 1–27.
46. Clark, S.; Coecke, B.; Sadrzadeh, M. A compositional distributional model of meaning. In Proceedings of the Second Quantum Interaction Symposium (QI-2008), Oxford, UK, 26–28 March 2008; pp. 133–140.
47. Buszkowski, W. Lambek grammars based on pregroups. In Proceedings of the International Conference on Logical Aspects of Computational Linguistics, Le Croisic, France, 27–29 June 2001; Springer: Berlin/Heidelberg, Germany, 2001; pp. 95–109.
48. Abramsky, S.; Coecke, B. Categorical quantum mechanics. *Handb. Quantum Log. Quantum Struct.* **2009**, *2*, 261–325.
49. Meichanetzidis, K.; Gogioso, S.; de Felice, G.; Chiappori, N.; Toumi, A.; Coecke, B. Quantum Natural Language Processing on Near-Term Quantum Computers. *Electron. Proc. Theor. Comput. Sci.* **2021**, *340*, 213–229. [[CrossRef](#)]
50. Pestun, V.; Vlassopoulos, Y. Tensor network language model. *arXiv* **2017**, arXiv:1710.10248.
51. Correia, A.; Moortgat, M.; Stoof, H. Grover’s Algorithm for Question Answering. *arXiv* **2021**, arXiv:2106.05299.
52. Abbaszade, M.; Salari, V.; Mousavi, S.S.; Zomorodi, M.; Zhou, X. Application of quantum natural language processing for language translation. *IEEE Access* **2021**, *9*, 130434–130448. [[CrossRef](#)]
53. Sordani, A.; Nie, J.Y.; Bengio, Y. Modeling term dependencies with quantum language models for ir. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 28 July–1 August 2013; pp. 653–662.
54. Xie, M.; Hou, Y.; Zhang, P.; Li, J.; Li, W.; Song, D. Modeling quantum entanglements in quantum language models. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
55. Li, Q.; Melucci, M.; Tiwari, P. Quantum language model-based query expansion. In Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval, Tianjin, China, 14–17 September 2018; pp. 183–186.
56. Jiang, Y.; Zhang, P.; Gao, H.; Song, D. A quantum interference inspired neural matching model for ad-hoc retrieval. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual, 25–30 July 2020; pp. 19–28.
57. Zhang, P.; Niu, J.; Su, Z.; Wang, B.; Ma, L.; Song, D. End-to-end quantum-like language models with application to question answering. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
58. Zhang, P.; Su, Z.; Zhang, L.; Wang, B.; Song, D. A quantum many-body wave function inspired language modeling approach. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 1303–1312.
59. Li, Q.; Wang, B.; Melucci, M. CNM: An Interpretable Complex-valued Network for Matching. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 4139–4148. [[CrossRef](#)]

60. Zhang, Y.; Song, D.; Li, X.; Zhang, P. Unsupervised sentiment analysis of twitter posts using density matrix representation. In Proceedings of the European Conference on Information Retrieval, Grenoble, France, 26–29 March 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 316–329.
61. Zhang, Y.; Li, Q.; Song, D.; Zhang, P.; Wang, P. Quantum-Inspired Interactive Networks for Conversational Sentiment Analysis. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, Macao, China, 10–16 August 2019; pp. 5436–5442. [[CrossRef](#)]
62. Zhang, P.; Zhang, J.; Ma, X.; Rao, S.; Tian, G.; Wang, J. TextTN: Probabilistic Encoding of Language on Tensor Network. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 4 May 2021.
63. Meichanetzidis, K.; Toumi, A.; de Felice, G.; Coecke, B. Grammar-Aware Question-Answering on Quantum Computers. *arXiv* **2020**, arXiv:2012.03756.
64. Lorenz, R.; Pearson, A.; Meichanetzidis, K.; Kartsaklis, D.; Coecke, B. Qnlp in practice: Running compositional models of meaning on a quantum computer. *arXiv* **2021**, arXiv:2102.12846.
65. Vicente Nieto, I. Towards Machine Translation with Quantum Computers. Master's Thesis, University of Stockholm, Stockholm, Sweden, 2021.
66. Mitchell, J.; Lapata, M. Vector-based models of semantic composition. In Proceedings of the ACL-08: HLT, Columbus, OH, USA, 15–20 June 2008; pp. 236–244.
67. Grefenstette, E.; Sadrzadeh, M. Experimental Support for a Categorical Compositional Distributional Model of Meaning. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Emnlp 2011, Scotland, UK, 27–31 July 2011.
68. BNC Consortium. British national corpus. In *Oxford Text Archive Core Collection*; University of Oxford: Oxford, UK, 2007.
69. Lewis, M. Towards logical negation for compositional distributional semantics. *IfCoLoG J. Logics Their Appl.* **2020**, *7*.
70. Tyrrell, B. Applying Distributional Compositional Categorical Models of Meaning to Language Translation. *Electron. Proc. Theor. Comput. Sci.* **2018**, *283*, 28–49. [[CrossRef](#)]
71. Sadrzadeh, M.; Clark, S.; Coecke, B. The Frobenius anatomy of word meanings I: Subject and object relative pronouns. *J. Log. Comput.* **2013**, *23*, 1293–1317. [[CrossRef](#)]
72. Coecke, B.; Pavlovic, D.; Vicary, J. A new description of orthogonal bases. *Math. Struct. Comput. Sci.* **2013**, *23*, 555–567. [[CrossRef](#)]
73. Kartsaklis, D.; Sadrzadeh, M. Prior disambiguation of word tensors for constructing sentence vectors. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1590–1601.
74. Schütze, H. Automatic word sense discrimination. *Comput. Linguist.* **1998**, *24*, 97–123.
75. Clark, S.; Coecke, B.; Grefenstette, E.; Pulman, S.G.; Sadrzadeh, M. A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure. *Malays. J. Math. Sci.* **2014**, *8*, 15–25.
76. Coecke, B.; Duncan, R. Interacting quantum observables: Categorical algebra and diagrammatics. *New J. Phys.* **2011**, *13*, 043016. [[CrossRef](#)]
77. Benedetti, M.; Lloyd, E.; Sack, S.; Fiorentini, M. Parameterized quantum circuits as machine learning models. *Quantum Sci. Technol.* **2019**, *4*, 043001. [[CrossRef](#)]
78. Hou, Y.; Zhao, X.; Song, D.; Li, W. Mining pure high-order word associations via information geometry for information retrieval. *ACM Trans. Inf. Syst. TOIS* **2013**, *31*, 1–32. [[CrossRef](#)]
79. Metzler, D.; Croft, W.B. A markov random field model for term dependencies. In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, 15–19 August 2005; pp. 472–479.
80. Zhang, P.; Li, J.; Wang, B.; Zhao, X.; Song, D.; Hou, Y.; Melucci, M. A quantum query expansion approach for session search. *Entropy* **2016**, *18*, 146. [[CrossRef](#)]
81. Wang, P.; Hou, Y.; Li, J.; Zhang, Y.; Song, D.; Li, W. A quasi-current representation for information needs inspired by Two-State Vector Formalism. *Phys. A Stat. Mech. Its Appl.* **2017**, *482*, 627–637. [[CrossRef](#)]
82. Robertson, S.E.; Walker, S. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In Proceedings of the SIGIR'94, Dublin, Ireland, 1 January 1994; Springer: Berlin/Heidelberg, Germany, 1994; pp. 232–241.
83. Van Rijsbergen, C.J. *The Geometry of Information Retrieval*; Cambridge University Press: Cambridge, UK, 2004.
84. Wang, M.; Smith, N.A.; Mitamura, T. What is the Jeopardy model? A quasi-synchronous grammar for QA. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague, Czech Republic, 28–30 June 2007; pp. 22–32.
85. Yang, Y.; Yih, W.t.; Meek, C. Wikiqa: A challenge dataset for open-domain question answering. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 2013–2018.
86. Yang, L.; Ai, Q.; Guo, J.; Croft, W.B. anmm: Ranking short answer texts with attention-based neural matching model. In Proceedings of the 25th ACM International Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; pp. 287–296.
87. Yin, W.; Schütze, H.; Xiang, B.; Zhou, B. Abcnn: Attention-based convolutional neural network for modeling sentence pairs. *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 259–272. [[CrossRef](#)]
88. Severyn, A.; Moschitti, A. Learning to rank short text pairs with convolutional deep neural networks. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 9–13 August 2015; pp. 373–382.

89. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics: Doha, Qatar, 2014; pp. 1746–1751. [[CrossRef](#)]
90. Poria, S.; Cambria, E.; Hazarika, D.; Majumder, N.; Zadeh, A.; Morency, L.P. Context-dependent sentiment analysis in user-generated videos. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 873–883.
91. Glasser, I.; Sweke, R.; Pancotti, N.; Eisert, J.; Cirac, I. Expressive power of tensor-network factorizations for probabilistic modeling. *Adv. Neural Inf. Process. Syst.* **2019**, *32*. [[CrossRef](#)]
92. Dai, B.; Li, J.; Xu, R. Multiple positional self-attention network for text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 7610–7617.
93. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.; Potts, C. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; Association for Computational Linguistics: Seattle, WA, USA, 2013; pp. 1631–1642.
94. Pang, B.; Lee, L. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04), Barcelona, Spain, 21–26 July 2004; pp. 271–278. [[CrossRef](#)]
95. Di Sipio, R.; Huang, J.H.; Chen, S.Y.C.; Mangini, S.; Worring, M. The Dawn of Quantum Natural Language Processing. *arXiv* **2021**, arXiv:2110.06510.
96. Basile, I.; Tamburini, F. Towards quantum language models. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7–11 September 2017; pp. 1840–1849.
97. Baez, J.; Stay, M. Physics, topology, logic and computation: A Rosetta Stone. In *New Structures for Physics*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 95–172.
98. Selinger, P. A survey of graphical languages for monoidal categories. In *New Structures for Physics*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 289–355.
99. Chomsky, N. Three models for the description of language. *IRE Trans. Inf. Theory* **1956**, *2*, 113–124. [[CrossRef](#)]
100. Spall, J.C. A one-measurement form of simultaneous perturbation stochastic approximation. *Automatica* **1997**, *33*, 109–112. [[CrossRef](#)]
101. Zhao, Q.; Hou, C.; Liu, C.; Zhang, P.; Xu, R. A quantum expectation value based language model with application to question answering. *Entropy* **2020**, *22*, 533. [[CrossRef](#)] [[PubMed](#)]
102. Chen, Y.; Pan, Y.; Dong, D. Quantum language model with entanglement embedding for question answering. *IEEE Trans. Cybern.* **2021**, 1–12. [[CrossRef](#)] [[PubMed](#)]
103. Sivarajah, S.; Dilkes, S.; Cowtan, A.; Simmons, W.; Edgington, A.; Duncan, R. $t|ket\rangle$: A retargetable compiler for NISQ devices. *Quantum Sci. Technol.* **2021**, *6*, 014003. [[CrossRef](#)]
104. Rimell, L.; Maillard, J.; Polajnar, T.; Clark, S. RELPRON: A relative clause evaluation data set for compositional distributional semantics. *Comput. Linguist.* **2016**, *42*, 661–701. [[CrossRef](#)]
105. Yoshikawa, M.; Noji, H.; Matsumoto, Y. A* CCG Parsing with a Supertag and Dependency Factored Model. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 277–287. [[CrossRef](#)]
106. de Felice, G.; Toumi, A.; Coecke, B. DisCoPy: Monoidal Categories in Python. In Proceedings of the 3rd Annual International Applied Category Theory Conference 2020, Cambridge, MA, USA, 6–10 July 2020; Electronic Proceedings in Theoretical Computer Science. Spivak, D.I., Vicary, J., Eds.; Open Publishing Association: The Hague, The Netherlands, 2021; Volume 333, pp. 183–197. [[CrossRef](#)]
107. Toumi, A.; Koziell-Pipe, A. Functorial Language Models. *arXiv* **2021**, arXiv:2103.14411.
108. Lambek, J. Pregroups and natural language processing. *Math. Intell.* **2006**, *28*, 41–48. [[CrossRef](#)]
109. Casadio, C.; Lambek, J. An algebraic analysis of clitic pronouns in Italian. In Proceedings of the International Conference on Logical Aspects of Computational Linguistics, Le Croisic, France, 27–29 June 2001; Springer: Berlin/Heidelberg, Germany, 2001; pp. 110–124.