A Quantum Theory of Terms and New Challenges to Meaning Representation of *Quanterms*

Diego A. Burgos
Wake Forest University
1834 Wake Forest Road, Winston-Salem, NC 27109
burgosda@wfu.edu

Abstract
This article discusses the challenges to meaning representation of terms posed by a quantum theory of terms (QTT) that was recently reported. We first summarize this theory and then highlight the difficulties of representing *quanterms*, which is the name we coined for the view that the QTT has of terms as quantum systems by analogy with quantum objects in quantum mechanics. We briefly summarize the representation practices followed to date to record and represent terminology. We use findings reported in the literature to model both terms and *quanterms* and found that current representations of terms in specialized repositories are collapsed *quanterms* at the expense of other states of the original *quanterm*. In this work, both *quanterms* and collapsed *quanterms* are mathematically modelled following formulations used in quantum mechanics. These formulations suggest that representations of *quanterms* need to include information about the probabilities of *quanterm* states and the role they play in the entanglement of terms for phenomena such as specialized collocations.

Keywords: terminology, quantum theory of terms, meaning representation

1. Introduction

In terminology, a term is operatively defined as a conventional, non-compositional lexical unit linked to a meaning exclusively used in a specialized domain, e.g., medicine, architecture, etc. (Burgos & Vásquez 2024). Traditionally, mainstream terminology theories and models define the term as a bidimensional object (see, for example, ISO 704, 2013, pp. 36-37; Cabré, 1999, p. 35; Faber and L'Homme, 2022, p. 355) with the term and a linked concept or meaning as the two dimensions of this representation.

However, Burgos et al. (2024) recently reported a quantum theory of terms (QTT), which models the term as a dynamic, multidimensional object with the characteristics of a quantum system. *Quanterms*, as they could be called, challenge the representation models that have been so far used to represent terms and their meanings. The implications of this quantum model may have a significant impact in computational linguistics, language engineering, lexicography and terminography, terminology theory and other fields related to knowledge representation, understanding and generation.

This paper highlights these challenges in the light of the QTT. In order to attain this, we summarize the most common representations of the term that have been used to date. Then, we briefly introduce the QTT as well as an abstract representation of *quanterms*. This background helps pave the way for a discussion section about the challenges of operative meaning representation of *quanterms*. We close with some conclusions and ideas for forms of representation.

2. Representation of terms

One of the most widespread representations of terms is the lexicographic representation, that is, the definition of terms in specialized dictionaries. Likewise, this representation has been the starting point of other forms of representation (e.g., Adelstein 2007, p. 72; Mahecha & De Cesaris 2011; Berri, 2013; Burgos & Vásquez 2024). For example, the lexicographic definition is frequently turned into Pustejovsky's generative lexicon model (1995, 2011), which, in turn, uses feature structures akin to those proposed by Carpenter (1992) to represent lexicon entries based on meaning features. These structures have also been utilized in other frameworks such as unification grammars (see, for example, Francez & Wintner, 2011) or semantic theories (e.g., naive semantics, Dahlgren, 1988). Naturally, terms also are represented in terminological databases generally following an onomasiological philosophy. This basically means that each term has one single sense and that each database entry or record hosts only one concept or sense together with the term or terms that denote it (cf. WordNet, Fellbaum 1998). Specialized taxonomies or ontologies such as SNOMED CT follow a similar approach.

According to Burgos et al. (2024), what these representations have in common is that they are static and limited, like pictures of a particular state of the term. While we acknowledge the importance of the role played by these representations throughout the history of knowledge management and representation, we believe that a quantum view of the term, which we summarize below, calls for representation of terms reflecting the complexity of quantum systems.

3. Quantum theory of terms and *quanterms*

Burgos et al. (2024) view the term as a complex, multidimensional object with dynamic properties. This complexity is the result of a number of states and dimensions, in which the same term exists simultaneously. At the moment of observation or measurement, the term collapses into a particular state and updates or *freezes* a set of its properties according to the collapsed state. We will see below that this collapse may also happen due to the term's

5th Workshop on Designing Meaning Representations (DMR 2024) @LREC-COLING-2024, pages 48–53
21 May, 2024.

interaction with its environment because of a quantum phenomenon known as *decoherence*. The property of having several states at the same time is called superposition, which is described below.

3.1 Superposition of terms

It is this complex nature described above that motivates Burgos et al.'s Quantum Theory of Terms (QTT) by analogy of terms with instances of quantum superposition. Superposition in quantum mechanics describes an object that has several different simultaneous states (Miret 2015, p. 83). In the medical domain, this superposition was exemplified by Burgos (2024) with two instances of a medical condition, which were given two distinct denominations, namely, *alien hand* and *anarchic hand*. These two terms turn out to be not just simple variants, but they seem to be motivated by two states of the term, each with its own configuration of features in the conceptualization of the syndrome at two different observation moments. Thus, the first state and its denomination reflect the sensation that the hand belongs to another person, while the latter indicates that the hand appears to refuse to obey its owner.

Additional evidence was reported by Burgos and Vásquez (2024) based on an experiment with a language model in the form of word embeddings also in the clinical domain in Spanish. They observed that, while *alteration* is the prototypical semantic class for the term *mutation* in specialized repositories, the data show semantic class variation for the same term in the same domain. Two additional semantic classes were detected, namely *entity* and *process*. Each of the contexts in which each variant of *mutation* occurs makes a distinct observation in the dimension of conceptual variation with effects on the term's properties. It is interesting to note that this variation may also impact the agency of the unit, i.e., whether *mutation* semantically acts as experimenter or agent.

One interesting trait of quantum superposition is that some of the possible states of a quantum system may be mutually exclusive. This happens, not only with the two perceptions of *alien hand* and *anarchic hand*, but also with the case of *mutation* above, since entities and processes are mutually exclusive. This non-coexistence of feature values has a significant impact in the way these terms are represented using, for example, a concept tree of the domain.

The quantum superposition of terms suggests the existence of basic conceptual variants, i.e., variants that do not change into another concept, but rather undergo a change in some of the features of the same concept. Using *mutation* as an example, and assuming we could map each of its states and assign its features a numerical value, we would have a first graphic model of term superposition, that is, three observations or states of *mutation* as a *quanterm*, which we illustrate in Figure 1.

The figure shows three different states of *mutation* on the *z*-axis where the values of features 1, 3, 4, and 6 (e.g., part of speech, predicativity, composition, and form) remain constant across states, but the values of features 2, 4, and 7 (e.g., class, agency, and function) change depending on the moment of observation. Visually, this variation can be seen as a change in the color tones for changing features compared to the uniform tones of the stable features. Theoretically, the term in isolation simultaneously has a number of states whose properties can only be determined at the moment of observation. Thus, the model in Figure 1 represents that three states of *mutation* coexist in the conceptual dimension and that, in each of these states, its class, agency, and function can change depending on the moment and dimension the *quanterm* is observed.

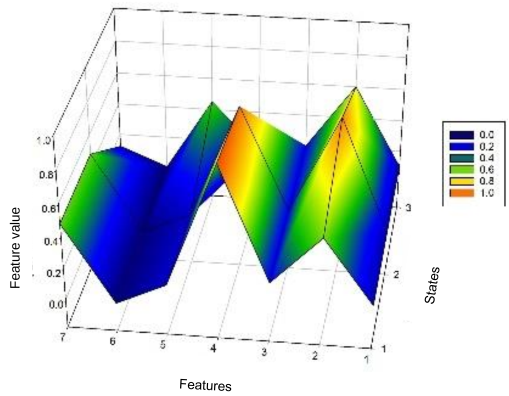

Figure 1. Model of *mutation* as a *quanterm* of three states

The model in Figure 1, however, can become more complex as the number of features, feature values, observations, and dimensions increase. The potential states of the *quanterm* in a more intricate scenario could therefore be represented by a matrix that combines these four factors. Regarding the possible number of dimensions, it is reasonable to think that it can always increase as more is known about the terminological phenomenon. However, we currently can predict six dimensions: dialect, level of specialization, social function, concept, domain, and time (see Table 1).

<table>
<tr><th colspan="2">Superposition dimensions</th></tr>
<tr><td rowspan="5">Time</td><td>Dialect</td></tr>
<tr><td>Specialization</td></tr>
<tr><td>Social function</td></tr>
<tr><td>Concept</td></tr>
<tr><td>Domain</td></tr>
</table>

Table 1: Dimensions where superposition can take place.

The reader may notice that these dimensions, except for *domain*, are related to a particular type of terminological variation reported in the literature (see Freixa, 2005). *Time* is an overarching dimension and accounts for diachronic aspects of the other five, including metaphorical phenomena. On the other hand, the *domain* dimension accounts for terms that can be at a crossroads between two or more domains or subdomains (e.g., *cell* in biology, veterinary science, and medicine). In each of these dimensions, the term can potentially take on a new feature or a different feature value in a particular observation.

As the *quanterm* becomes more complex, Figure 2 attempts to represent multiple states and features of a hypothetical *quanterm* in multiple dimensions.

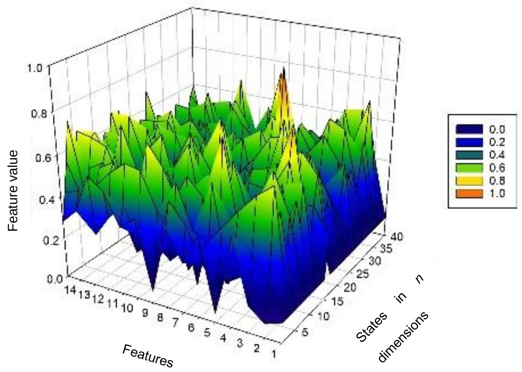

Figure 2. Hypothetical *quanterm*

For the model in Figure 2, we use a normal distribution of hypothetical feature values. It is a conservative representation based on the assumption that the *quanterm* is reasonably stable even though its features can be variable; otherwise, it would end up being a different concept. Visually, this stability can be seen as a lot of green and blue in the middle area of the graph with some peaks of color variations for significant changes in feature values that occur in particular states.

An example of the potential growth in the number features of a *quanterm* can be seen in *palatine tonsil*. The fact that that this term can be defined from a number of different subdisciplines increases the number of features that make up this *quanterm*. Depending on when and where the measurement of this term happens, its feature configuration would change. This occurs because a dentist, for example, gives prominence to features that may not be relevant to an anatomist, a pathologist, or a speech therapist, who would in turn highlight other features of the term when they use it in their respective domains while being the same *quanterm*.

A caveat is necessary here that the model in Figure 2 does not capture yet another layer of complexity added by the interaction or interdependence between *quanterms*. We describe such interaction below, which the QTT calls term entanglement.

3.2 Term entanglement

In quantum mechanics, entanglement refers to the interaction between particles such that the state of an object can be used to predict the state of another object (Miret 2015, p. 126). This property of *quanterms* allows for measuring the state of one term anticipating at the same time information about the state of other terms. A hypothetical example of this interaction can be the impact that a variation in the semantic class of a term in the conceptual dimension may have on, say, the agency of another term in the same dimension or in a different one. The QTT predicts that entanglement can happen even if the involved terms are far away from each other.

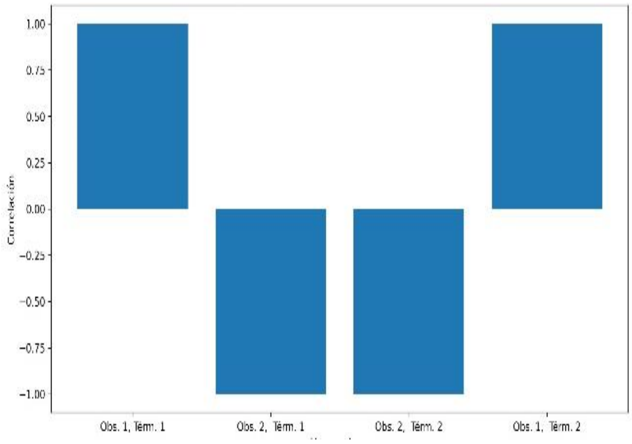

Figure 3. Entanglement of *quanterms*

Figure 3 illustrates possible effects of entanglement between two *quanterms*. The first observation of term 1 (Obs. 1, Term. 1) predicts that the state of Term 2 (Obs. 1, Term. 2) changes in a positive correlation. That is, if we assign numerical values to the features of Term 1, the values of Term 2 would change in the same direction. In the second observation of the same term, however, the correlation is negative. In other words, if the values of Term 1 increase, those of the other one decrease.

Besides other phenomena, entanglement of *quanterms* can explain specialized collocations. For example, a predicative term like *cancer* often selects terms referring to organs or tissues, such as *breast*, *prostate*, *stomach*, *skin*, etc., to produce collocations such as *breast cancer*, *gastric cancer*, *prostate cancer*, etc. Entanglement allows for predicting a correlation between the feature values of *cancer* and those of its collocation bases. Thus, we could reasonably anticipate that if *cancer* has a high value for the feature *alteration* (i.e., disease), the value of the feature *disease target* will proportionally increase in terms like *prostate* or *breast*. Let us use the definition of *prostate* in the Mosby Medical Dictionary (Villanueva et al. 1999) to clarify this point. *Prostate* is defined and anatomically described as a male gland, but there is no feature in its definition indicating that this gland is a target of cancer, perhaps because its value in this measurement of the term is very low. The feature *disease target*, however, is activated or its value increases in the domain dimension of oncology in positive correlation with the value of the feature *alteration* of terms like *cancer*. This entanglement, though, may not occur to the same extent for *prostate* in other domains such as anatomy or urology, which seems to be the state defined by the Mosby Dictionary above. The QTT attributes particular definitions or conceptualizations of terms to quantum decoherence, a phenomenon that is described below.

3.3 Decoherence

There are two reasons why a quantum system may collapse into one of its states, namely, the mere act of measuring it and its interaction with its environment. In quantum mechanics, this collapse is known as decoherence. The term is considered a *quanterm*, i.e., a quantum system, because, in isolation, it is in an indetermined number of different states at the same time, that is, it can be defined in multiple ways, even though some of those definitions may seem to conflict with each other. Any term in abstract, without any further definition or textual context, is a *quanterm*. Measuring a term may take the form of defining it or conceptualizing it, which involves determining the semantic features that delimit its specialized meaning. When this measuring operation takes place, the multiple states of the *quanterm* collapse into the meaning or conceptualization that it has been given, and it becomes a *collapsed term*, like the ones we currently see represented in dictionaries or ontologies.

Decoherence also happens as soon as the *quanterm* interacts with its environment, that is, with other terms and expressions in the context of a specialized text. The more specialized and specific the context, the more delimited its state is. We have an example of this explained above; the *quanterm* "*mutation"* collapses into a different state (entity, alteration, or process) depending on what environment it interacts with. It must be considered also that decoherence may sometimes be conditioned by a term entanglement.

4. *Quanterm* representation challenges

It is important to clarify that *quanterm* superposition does not refer to polysemy, but to the same term, concept or sense, which undergoes at a higher level a number of states (i.e., variations) at the same time, even if they are mutually exclusive, without changing into another concept. Polysemy has been successfully handled by lexicographic representations with a semasiological orientation (i.e., general dictionaries) and semantic networks, such as WordNet, as well as by formalisms based on qualia structures, such as the one in Figure 4 reported by Núñez Torres (2013).

[Door (x ∨ y)
QUASTR [FORM: physical_object´ (x), frame´ (y)
CONST: obstruction´ (x), aperture´ (y)
TELIC: BECOME closed´ / open´ (x), do´ (z, [go.through´ (z, y)])
AGENT: artifact´ (x ∨ y)]]

Figure 4. Representation of a collapsed state of the polysemous term *door*

Of a similar nature are onomasiological specialized resources such as terminological databases and ontologies. All these resources, however, always record collapsed states of *quanterms*. See, for example, the representation of the collapsed concept *myocardial infarction* in SNOMED CT in Figure 5[1], which, interestingly, seems to be also an instance of term entanglement, also known as a specialized collocation.

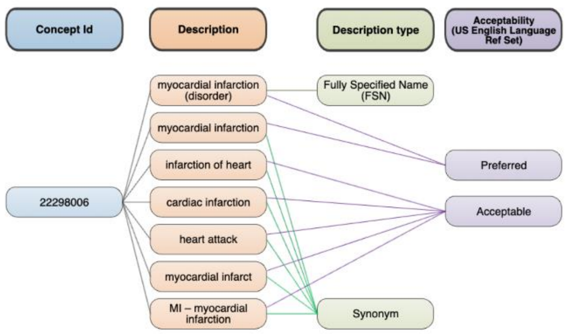

Figure 5. Representation of a collapsed state of *myocardial infarction*

The extended practice of representing terms in a collapsed form may be due to the difficulty of representing more complex systems, but the likely reason for this appears to be that terms had not been seen before as the quantum systems proposed by the QTT. Due to their complexity, the representation of quantum systems is generally mathematical. The mathematical formulation of a quantum system is independent of the type of system; therefore, we can represent a *quanterm* of an undetermined number of states with the equation in Figure 6:

$$|\Psi(t)\rangle = \sum_n C_n(t)|\Phi_n\rangle$$

Figure 6. Mathematical formulation of a *quanterm*

The psi symbol at the left of the equation is the conventional notation for a system in superposition, that is, a *quanterm* in our case. It is equal to the sum of the amplitude probabilities of observing particular states, where *n* stands for the number of states of the system.

[1] SNOMED CT Starter Guide at https://confluence.ihtsdotools.org/display/DOCSTART

In contrast, the mathematical formulation of collapsed *quanterm* representations, such as the ones in Figures 4 and 5, is simpler:

$$P_j = |a_j\rangle\langle a_j|$$

Figure 7. Mathematical formulation of a collapsed *quanterm*

This equation uses the projection operator P_j to show the projection of the quantum system onto the state $|a_j\rangle$ associated with the measurement outcome a_j (i.e., a definition, conceptualization, feature structure, etc.).

The main challenges related to the QTT are, then, *1)* to represent *quanterms* either by using traditional formats, which seem to be limited for this purpose, or by innovating more sophisticated formats and *2)* to take advantage of such representations to use the potential of *quanterms* for faster and more efficient and intelligent language processing tasks. In quantum mechanics, an electromagnetic wave is sent to the quantum object to verify superposition and to learn about the potential states of the object. According to the equation in Figure 6, the key knowledge learned seems to be the probability of a state happening at a given observation of the object. In times of deep learning and artificial intelligence, language models may play the role of this wave to determine such probabilities.

The representation of a *quanterm* like *mutation* should include, then, the probabilities to predict not only its potential classes but also other variations in its features. These probabilities will make even more sense if they are conditioned by and linked to any relevant entanglement with other *quanterms* and with its context itself.

On the other hand, the potential combination of efficient representation of *quanterms* with modern supercomputing may be necessary. The optimal utilization of *quanterms* and their representation may add to the newly born quantum semantics landscape (see an example of a work that attempts term entanglement in Surov et al., 2021).

5. Conclusions

This paper presents some of the challenges of meaning representation of terms in the light of a quantum theory of terms (QTT) recently reported by Burgos et al. (2024). Due to the novelty of the QTT, we summarized the theory and coined the expression *quanterm* to denote terms viewed as quantum systems. Our focus in this work, however, was on the limitations of representation forms traditionally used in terminology and on the need for innovative representations to respond to the nature of *quanterms*. We highlighted that those traditional representations of terms actually record collapsed *quanterms* at the expense of other potential states (i.e., conceptual variations) of the documented terms.

A comparison of the mathematical formulation of *quanterms* versus collapsed *quanterms* showed the complexity that is being lost in current representations. It was noted that the probabilities to predict particular states of a quantum system are key, not only to this mathematical formulation, but also to potential envisioned forms of *quanterm* representations. Finally, term superposition and entanglement may play an important role not only in term extraction and collocation identification but also in text categorization and knowledge representation, understanding, and generation.

6. Bibliographical References

Adelstein, A. (2007). *Unidad léxica y significado especializado: modelo de representación a partir del nombre relacional madre* [Tesis doctoral, Universidad Pompeu Fabra].

Berri, M. (2013). Léxico generativo y aplicaciones lexicográficas: Los nombres concretos del dominio de la medicina en el DRAE. *Revista signos*, 46(82), 190–212.

Burgos, D. & Vásquez, D. (2024). El nombre terminológico. En G. Quiroz, D. Burgos, & F. Zuluaga (Eds.), *Terminología del español: el término*. Routledge.

Burgos, D. (2024). El término complejo. En G. Quiroz, D. Burgos, & F. Zuluaga (Eds.), *Terminología del español: el término*. Routledge.

Burgos, D., Quiroz, G. & Pérez-Pérez, C.M. (2024). Antecedentes y principios para una teoría cuántica del término. En G. Quiroz, D. Burgos, & F. Zuluaga (Eds.), *Terminología del español: el término*. Routledge.

Cabré, T. (1999). Hacia una aproximación teórica de base comunicativa. Elementos para una teoría de la terminología: hacia un paradigma alternativo. In M. Cabré (Ed.), *La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos* (pp. 69–92). Instituto Universitario de Lingüística Aplicada.

Carpenter, B. (1992). The logic of typed feature structures: Applications to unification grammars, logic programs, and constraint resolution. *Cambridge tracts in theoretical computer science*. Cambridge University Press.

Dahlgren, K. (1988). *Naive semantics for natural language understanding*. Kluwer Academic Publishers.

Faber, P., y L'Homme, M. (2022). *Theoretical perspectives on terminology: Explaining terms, concepts, and specialized knowledge*. John Benjamins Publishing Company.

Fellbaum, C. (1998). WordNet: An electronic lexical database. MIT Press.

Francez, N., y Wintner, S. (2011). *Unification grammars*. Cambridge University Press.

Freixa, J. (2005). Variación terminológica: ¿por qué y para qué? *Meta*, 50(4), https://doi.org/10.7202/019917ar

Instituto Colombiano de Normalización (Icontec) e International Organization for Standarization (ISO). (2013). *Trabajo terminológico. Principios y métodos* (NTC-ISO 704).

Mahecha, V., y DeCesaris, J. (2011). Representing Nouns in the Diccionario de aprendizaje del español como lengua extranjera (DAELE). In I. Kosem y K. Kosem (Coord.), *Electronic lexicography in the 21st century: New applications for new users: Proceedings of eLex* (pp. 180–186). Bled.

Miret, S. (2015). Mecánica cuántica. CSIC – Consejo Superior de Investigaciones Científicas.

Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.

Pustejovsky, J. (2011). Coercion in a general theory of argument selection. *Linguistics*, 49(6), 1401–1431.

Surov, I. A., Semenenko, E., Platonov, A. V., Bessmertny, I. A., Galofaro, F., Toffano, Z., ... & Alodjants, A. P. (2021). Quantum semantics of text perception. *Scientific Reports*, 11(1), 4193.

Torres, F. N. (2013). La representación léxica en el modelo del Lexicón Generativo de James Pustejovsky. *Onomázein*, (28), 337-345.

Villanueva, A., López, C., y Ruiz, A. (1999). *Diccionario Mosby: de medicina, enfermería y ciencias de la salud*. Mosby.