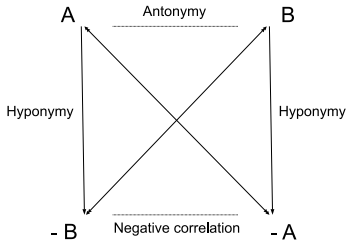
Entanglement and Semantics: Application to Language Processing

Francesco Galofaro, Zeno Toffano, Bich-Liên Doan, Adam Shimi

Entanglement and Semantics : Application to Language Processing

Francesco Galofaro[1] Zeno Toffano[2,4] Bich-Liên Doan[3,4] Adam Shimi[3,4]
[1]*Politecnico, Via Durando 38/A, 20158, Milan, Italy.* [2]*Laboratoire de Signaux et Systèmes (L2S, UMR8506)* [3]*Laboratoire de Recherche en Informatique (LRI, UMR8623) - CNRS - Université Paris Saclay* [4]*CentraleSupélec, 3 Rue Joliot Curie, 91192 Gif-sur-Yvette, France*

Entangled Bell states have been represented in quantum computation by a $CNOT$ gate whose control Qbit is in an undetermined state. This basic circuit can be useful to represent semantic relationships between two words (lexemes) in a text. Entangled states have been used to detect semantic relationships in Information Retrieval (IR). A lexeme can be represented as a vector in a Hilbert space [1]. Each element of the vector represents the weight of a semantic relation with a second vector : the context of a second lexeme. Two query operators are defined in a way that they attribute the value $+1$ to the component of the state that corresponds to the word meaning we are interested in, and -1 in the orthogonal direction : they correspond to Pauli spin matrices $\widehat{\sigma}_x$ and $\widehat{\sigma}_z$ [4]. In particular, since the first operator corresponds to negation and the conditional gate corresponds to an implication in classical logic, they can represent two basic semantic relationships : hyponymy (e.g. genus-species) and antonymy (e.g. masculine-feminine). In fact, according to Semantics, if *A* and *B* are antonymic terms, then *A* is an hyponym of (i.e. implies) *not-B* [5] as in the semiotic square [6] :



For example, if we consider the New Testament, Matthew, 7, 24-27, "sand" is antonymically opposed to "rock" as it concerns the groundworks of houses. We can represent the relation of entanglement between "rock" and "house" as a $CNOT$ gate between two Qbits. The control Qbit represents the presence/absence of the lexeme "rock". A similar relation of entanglement can represent the relation between "sand" and "house". The antonymy relation can be described by the Z gate. The control Qbits will be in an undetermined state, whose amplitude of probability can represent the strength of the semantic relation, which depends on the distance between the two lexemes in the text according to IR models [4]. We underline how the relations of a given semantic universe, and the corresponding result of our operators, depends only on the considered text : for example, wikipedia defines "sand" as an hyponym of "rock". In terms of structural semantics, entanglement represents the presence/absence of contextual semantic relations (*classemes*). The reason consists in the fact that linguistic categories are relational : for example, a "subject" exists only in opposition to an "object" and does not show any independent, positive identity [3]. Among others [2], this analogy suggests the application of quantum-based models to language and IR.

[1] K. Van Rijsbergen, The Geometry of Information Retrieval, Cambridge : Cambridge University Press (2004).
[2] Y. Lee, H. Cunningham, "Geometric and Quantum Methods for Information Retrieval", ACM SIGIR Forum, **Vol. 42 No. 2**, 22-32 (2008).
[3] F. Galofaro, "Structural Syntax and Quantum Computation" in Morphogenesis and Individuation, Berlin : Springer, 173-201 (2014).
[4] J. Barros, Z. Toffano, Y. Meguebli, and B. Doan. "Contextual Query Using Bell Tests", QI 2013 LNCS **8369**, 110-121 (2014).
[5] J. Lyons, Semantics 2, Cambridge : Cambridge University Press (1977).
[6] A.J. Greimas and F. Rastier, "The Interaction of Semiotic Constraints", Yale French Studies, **41**, 86-105 (1968).