

## Semantic Quantum Correlations in Hate Speeches

**Francesco Galofaro**

Università degli Studi di Torino - DFE Dipartimento di Filosofia e scienze dell'educazione  
francesco.galofaro@unito.it

**Zeno Toffano**

CentraleSupélec, Laboratoire des signaux et systems  
zeno.toffano@centralesupelec.fr

**Bich Lien Doan**

CentraleSupélec, Laboratoire des signaux et systems  
bich-lien.doan@centralesupelec.fr

**Abstract** The intervention shows the first results of a research conducted on a corpus of 7000 posts collected on the Reddit social network during the 2016 American presidential campaign. The research is the result of a collaboration between Berkeley D-Lab, who shared the corpus, LSI - CentraleSupélec and CUBE. Thanks to funding from the Anti-Defamation League, the corpus has been labeled to apply Machine Learning techniques: 400 posts have been labeled as “hate speech” by human analysts. Galofaro, Toffano and Doan applied to both sub-corpora (hate and non-hate speeches) an analysis technique inspired by Greimas’s structural semantics, Eco’s semiotics, and Quantum Information Retrieval (van Rijsbergen).

Each text was formalized as a semantic network using the HAL technique. We then measured the semantic similarity between two key words formalized as two word-vectors with the classical measure of cosine-similarity and then compared it with the degree of quantum correlation between them measured with the Born rule. This correlation, linked to the co-occurrence of the word vectors in the same contexts, extracts from the latter useful information to characterize the considered semantic relationships (“presence of correlation”, “absence of correlation” or “presence of anti-correlation”). In this way, the new technique allows to overcome some critical aspects of the Machine Learning techniques currently in use, being based on the meaning of the text and not on the way in which the human analyst labels the corpus.

**Keywords:** Semiotics, Semantics, Quantum Information Retrieval, Hate Speech, Political discourse

Accepted 8 May 2020.

## **1. Hate speeches: definition and problems**

According to John Nockleby hate speech is «any communication that disparages a person or a group on the basis of some characteristics (to be referred to as types of hate or hate classes) such as race, colour, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics» (Nockleby 2000). Hate speeches became a political problem in parallel with the diffusion of social networks. In 2012 The percentage of European young people which have encountered hate speech online stood at 80%, while the percentage of young people which felt attacked or threatened stood at 40% (cf. Aa.Vv. 2012). However, considering the lack of linguistic features considered by current definitions of ‘hate speech’, it is difficult identify them. Automatic detection and censure of hate speeches is a sensible problem in relation to freedom of speech. The vulgar and offensive meaning of hate speeches is not related to a referent, ontologically located in the word and objectively identifiable by the participants to the communicative process; hate speeches rather involve enunciation, and, in particular, the subjectivity of the receiver.

### **1.1 The corpus**

A possible solution is represented by the application of statistical methods to let emerge the features of hate speeches directly from a corpus of messages. To this purpose, a corpus has been collected by Berkeley D-Lab, thanks to the funding of the Anti-Defamation League. The corpus counts 7619 posts on the social platform Reddit dating back to the US Presidential Elections of 2016. The goal is to apply Machine Learning techniques to this corpus, in order to recognize hate speeches without having to specify their linguistic features. The corpus has been labeled by humans (trained students), and 411 texts have been considered ‘hate speeches’. The top 5 words used in the hate-speeches subset are: Jews, White, Hate, Black, Women. Among these, white and black are interesting because they can be considered an antonymic couple from the point of view of lexical semantics.

### **1.2 Problems**

The increasingly widespread use of neural networks and machine learning techniques in the legal field raises ethical questions. It is a commonplace that machines are immune from human biases; on the contrary, machines absorb biases from their corpus. Thus, human responsibility is always questioned, as well as the possibility of manipulating algorithms to reach ideological goals, presenting the decision of the machine as ‘objective’, using it to limit freedom and to delegitimise the political opponent’s point of view.

A second threat is represented by ‘ethical outsourcing’. A characteristic of our time is to delegate philosophy to machines. In fact, automatic ethical judgment is only the final step after the success of aesthetic and ontological algorithms:

- we ask search engines to measure the relevance of images to our queries;
- we ask algorithms to report fake news: European Union financed a project (<https://askpinocchio.com/>) that claims to assign a probability value to news on the basis of a textual analysis, thus confusing the credibility of the lexicon with the reference to a state of affairs.

However, when one asks if it is right to entrust moral judgment to Artificial Intelligence, the problem is whether a not-human, artificially created ‘intelligence’ exists or not. If we paraphrase the question into ‘it is right to entrust moral judgment to a new statistical approach’, the debate would gain in clarity.

### 1.3 Technical limitations

Finally, and most importantly, the actual automatic classification techniques based on neural networks show important technical limitations. Neural Networks are less efficient when the goal is to distinguish between a great number of classes; as a consequence, Neural Networks find it hard to classify hate speeches in genres. In fact, hateful content lacks of unequivocal linguistic features (Zhang & Luo 2018). As we said, ‘hate’ involves the intersubjective dimension of enunciation: hate speeches show a philosophical relevance.

This suggests a closer analysis to the *immanent* semantic features of the hate speech. On this purpose, it is necessary to adopt a different technique, inspired to quantum geometry, that will be presented in paragraph 2.

## 2. A quantum semantic memory

In the proposed model, the semantic dimension of the document will be considered as a quantum semantic memory (QSM), which can be retrieved and modified by a quantum logic unit (QLU). The QSM is a net of context-sensitive relations between lexemes. These relations are weighted, and depend mainly on the distance between the two considered lexemes. They are re-enforced whenever two lexemes co-occur more than one time in the text. The weight of a relation plays the role of probability in the quantum formalism. From this point of view, the text does not appear any more as a discrete net of words, but as a geometrical space (the semantic space). The QLU is the algorithm we use to transform the space in order to let emerge the stronger and the weaker relations we are interested in. The QLU consists of *operators*, acting on the semantic relations. The nature of these operators is *logical*, with particular reference to quantum logic (see Dubois and Toffano 2017).

This model is consistent with the semiotic tradition. The notion of semantic memory has been proposed firstly in the seminal works by Ross Quillian (1968), and it is the basis of Umberto Eco’s notion of encyclopedic format (cf. *Semantica della metafora*, in Eco 1972). Quillian’s model is deterministic, while posterior research turned its attention towards probability, understood as a measure of the weight of semantic relations. The present research is based on the model of semantic memory proposed by Lund and Burgess (1996).

A detailed technical exposition of the quantum formalism applied to hate speeches has been presented in Galofaro, Toffano, and Doan (2018). Here we will focus on semiotics, in particular on structural semantics. We will describe step-by-step how our algorithm works on a text belonging to our corpus:

That’s probably because 30 years ago they were not bashing black or women. Well, women only got bashed if they mouthed off.

In our corpus, this text has been labeled as non-hate speech. It contains the words black (B) and women (W), and it does not contain the word ‘white’.

### 2.2 From words to lexemes

The first step is to convert the text in a quantum semantic memory, using the HAL method (cf. Lund and Burgess 1996). The method consists in producing a matrix whose rows and columns represent the lexemes occurring in the text. We need to obtain the lexemes from the words both for semantic and technical reasons: we want to avoid that the computer considers the singular and the plural of a lexeme as two distinct words. To

obtain this we apply a *stemmer*, a standard library of the python programming language capable of reducing each word to a specific stem, similar to – but not coincident with – the linguistic notion of root (e.g. black, blacks, blackness)<sup>1</sup>. There is a risk of oversimplification, but every model has to renounce, in principle, to some information to focus on structural phenomena.

### 2.3 The quantum semantic memory

We also set an optimal window, i.e. the length of the context we want to consider (window). In the considered example, we consider a window of 11 word. By moving the window lexeme by lexeme over the document we write in each square of the matrix a number which is inversely proportional to the distance between the lexeme-row and the lexeme-column. We finally sum the different occurrences of a lexeme: for example, women occurs two times in our document. We can see the result in fig. 1:

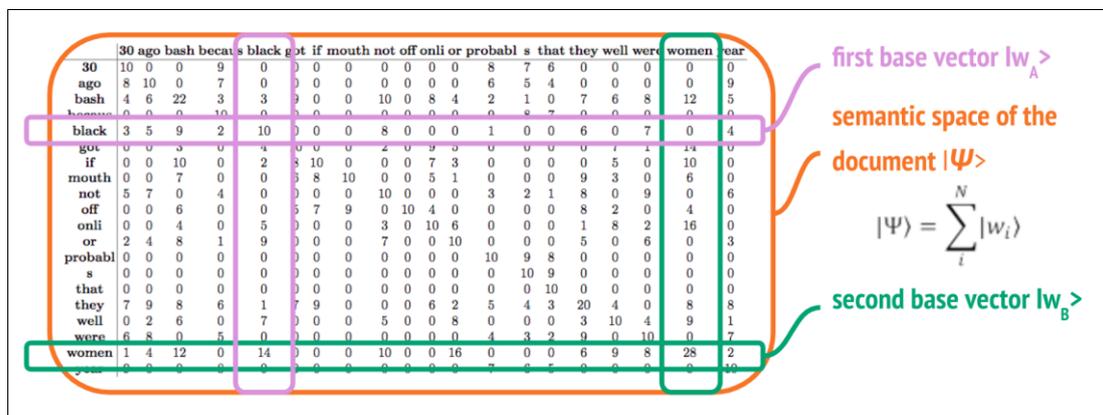


Fig. 1 - A Quantum Semantic Memory

Fig. 1 represents the semantic space of the document. Each lexeme is represented as a row and a column vector. In the figure we underlined two word-vectors (women and black). The vectors represents the relations between these two lexemes and all the other lexemes of the document, in each context provided by the document. They tell us something (information) on the distribution of their meaning along the textual space. In other terms, each of them represents an *isotopy*, defined as coherent semantic layers (cf. Greimas 1966). As Guido Ferraro notices (2019: 66) the isotopic effect derives from the structural oppositions on which it is actually based: every narration needs to oppose values. For example, in a first document, black women can be considered the opposite of white women; in a second speech black women can be considered as a subset of black people. Thus, the semiotic square provides the basic oppositions we can find between two textual isotopies: contradiction, implication, antonymy, sub-contrariety – see Greimas and Rastier (1968). However, since we are interested in the isotopic dissemination starting from any two lexemes, we are interesting in *measuring the strength of the considered opposition*.

### 2.4 The big question

How to acquire information on the relation between the “black”-isotopy and on the “women”-isotopy in the semantic space? First, we want to know whether they are related or not. But, in a document all meanings are related. Thus, we are interested in

<sup>1</sup> After different attempts we opted for Lancaster stemmer. Less aggressive libraries such as the Porter stemmer still distinguish between singular and plural. We also eliminated every information manifested by morphology, syntax, and punctuation.

the weight of this relation. Second, we are interested in the *type* of semantic relation between the two isotopies: are ‘black’ and ‘women’ opposed, as they were antonyms? Does the text give them a similar meaning, as they were synonyms? Does the first presuppose the second (or vice-versa)? Finally: where to find information to typify the semantic relation?

The use of the term ‘antonym’ we made above might leave puzzled: ‘black’ and ‘women’ are not registered as antonyms in the dictionary. As we will make it clear, the text constructs their opposition. This has been explained by Rastier (2009) as a transfer of semantic values not belonging to the functional system of the language, but to other systems, such as social or idiolectal norms (*afferents semes*). In our case, these values are proper to specific political and sub-cultural groups.

### 2.5 Geometric transformations

The procedure to convert the information of the the semantic memory in a more comfortable-to-retrieve format is reported in Galofaro, Toffano, and Doan (2018). Here we will focus only on those features which seem more relevant to semantics. In fig.1 we reported a formula that allows us to transform the semantic space in a single document vector,  $|\Psi\rangle$ . This vector, which represents the sum of all the isotopies, can be expressed in different bases: in particular, we can choose the two lexemes we are interested in as a base (fig. 2).

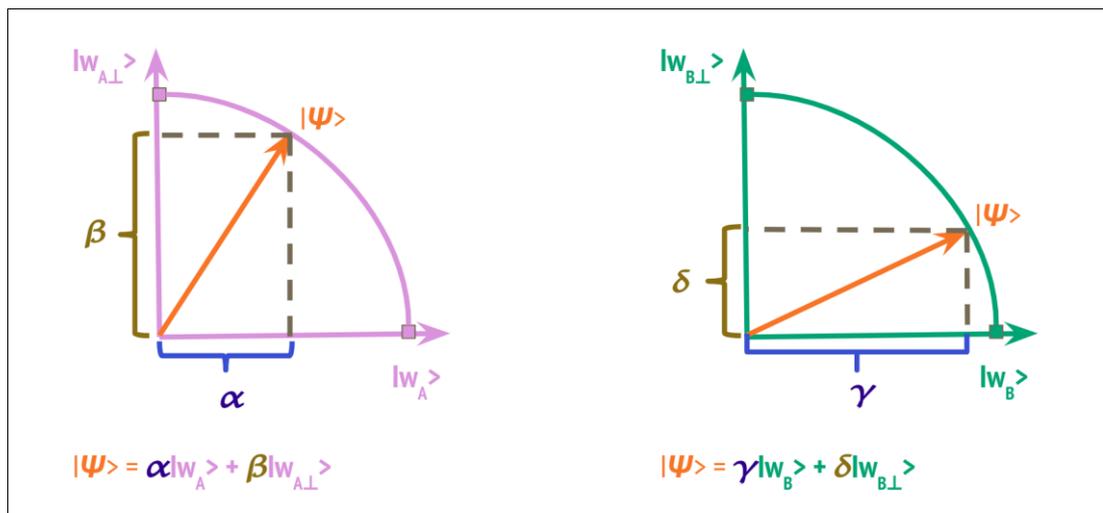


Fig. 2 – The same document  $|\Psi\rangle$  can be expressed in the two different bases provided by the keywords we are interested in (women Vs. black)

In fig. 2 we can see the same document-vector ( $|\Psi\rangle$ ) expressed in terms of its respective projections on two different bases by the theorem of Pythagoras. The first base is provided by the word-vector ‘black’ ( $|w_A\rangle$ ) and by its orthogonal vector ( $|w_{A\perp}\rangle$ ). the second one is provided by the word-vector ‘woman’ ( $|w_B\rangle$ ) and by its orthogonal vector.

### 2.6 Semantic interpretation of orthogonality

It has to be noticed how, when the ‘black’ component is at the maximum (when the  $|\Psi\rangle$  vector is parallel to the  $|w_A\rangle$  base), the value of the projection on ( $|w_{A\perp}\rangle$ ) is 0 and vice versa. The same can be said about the base provided by the word-vector ‘woman’ and its orthogonal vector. Thus, we can interpret the orthogonal vector as ‘absence’ of semantic value (respectively: absence of ‘woman’, absence of ‘black’). This

is consistent with Greimas definition of contradiction: the presence of one term presupposes the absence of the other and vice-versa – see ‘contradiction’, ‘semiotic square’ in Greimas and Courtés (1979).

### 3. The Quantum Logic Unit

To retrieve the Quantum Semantic Memory we need a Quantum Logic Unit: a set of operators capable of transforming meaning. The next step will be:

- to transform the meaning of the document expressed on the black-base ( $|w_A\rangle$ );
- to transform the meaning of the document expressed on the women-base ( $|w_B\rangle$ );
- to measure the expected outcome when the two transformations are applied together;

To construct our operators, we choose the X gate in quantum computation and we define:

- the  $B_x$  operator, which inverts the black-related meanings in the document vector;
- the  $W_x$  operator, which inverts the women-related meanings in the document vector;

For example, in fig. 3, we represent how the  $B_x$  operator transforms the document-vector, switching the  $\alpha$  and the  $\beta$  component.

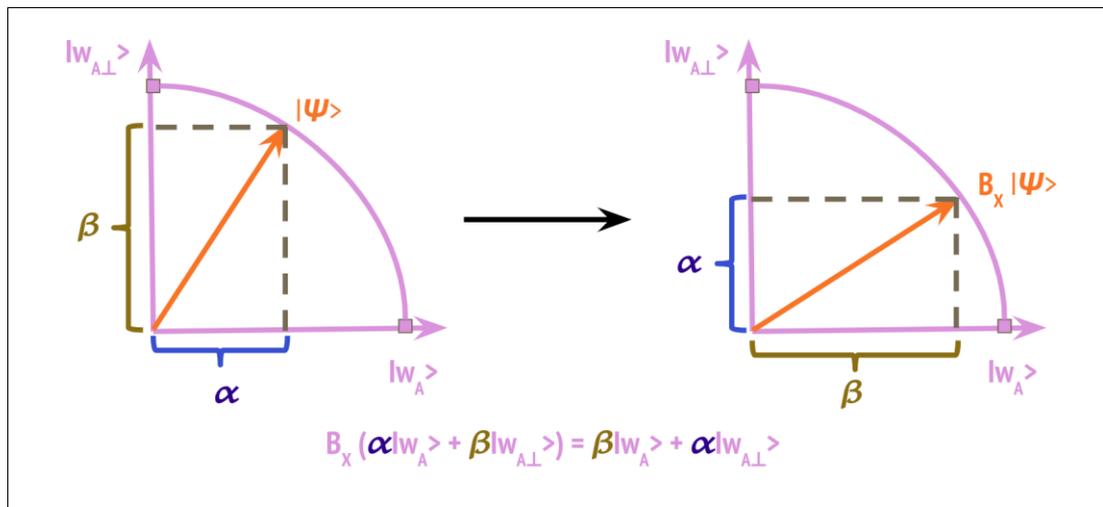


Fig. 3 - How the X operator transforms all the semantic values associated to the black-lexeme by rotating the document vector

This is consistent with semiotic notion of meaning as transformation and of theory as the rules of controlled transformations:

The construction of this space that we need will therefore coincide with the theory itself, that is to say with all the constituent categories that are organized in a structured system. Here, the structure is above all the organization of the conditions of possibility of the phenomena, but it is revealed immediately [...] as the scientific form of their description, the controlled form, by inter-definition, of the necessary practice (and thus universal) which consists in paraphrasing, repeating, transforming the given meaning into a new meaning (Marsciani 2014).

### 3.1 Expectations

What happens when we apply the two operators on the document at the same time? There are three interesting scenarios:

- 1) Every time the first operator changes a lexeme (+1) the second operator changes the same lexeme (+1). Every time the first machine leaves unchanged a lexeme (-1) the second machine leaves it unchanged (-1). If we multiply the two outcomes (+1,+1) or (-1,-1) we have an expectation value of +1: the two meanings /black/ and /women/ are correlated in the document.
- 2) Every time the first machine changes a lexeme (+1) the second machine leaves it lexeme unchanged (-1). Every time the first machine leaves unchanged a lexeme (-1) the second machine changes it (+1). If we multiply the two outcomes (+1,-1) or (-1,+1) we have an expectation value of -1: the two meanings /black/ and /women/ are anti-correlated in the document.
- 3) The changes can be concomitant in some context while in others they are not concomitant not (+1,+1); (+1,-1); (-1,+1). Their average is (0). Interpretation: the the two meanings /black/ and /women/ are not correlated in the document.

Obviously, all the values between -1 and +1 are a measure of a stronger or a weaker semantic (anti-)correlation. The expectation value is helpful to typify the semantic relation we are interested in. It is possible to calculate it applying the Born rule:

$$\langle \Psi | B_x W_x | \Psi \rangle$$

### 3.2 Bell Value

Beside  $B_x$  and  $W_x$  it is possible to define other operators starting from Pauli gates in quantum computation. In particular, we are interested in Pauli Z-gate, since, with an opportune choice of the operators, it is possible to calculate the Bell value (S). The formula is presented and commented in Barros, Toffano, Meguebli, and Doan (2014). As in quantum theory, the Bell value is less than or equal to 2 when there is a classical correlation between the two lexemes; it is in the range of 2 to  $2\sqrt{2}$  (approximately 2,8) when there is a quantum correlation.

In Quantum Information Retrieval, the variation of this value in relation to the considered context has been considered a measure of the semantic relation between two queries A and B ('A in the sense of B'). Here we are going to consider a fixed window and to compare the expectation value to the Bell value. Since we empirically measure both classical and quantum correlation it becomes critical to provide a semantic interpretation to the difference between the classical and the quantum correlation. We will try to do that basing on our corpus of hate speeches.

## 4. Findings

As we wrote above, the correlation value allows us to typify the semantic correlation between the two isotopies we are interested in, whereas the Bell value allows us to distinguish between classical and quantum correlations. Basing on these two values, we can distinguish four kinds of relations between isotopies in the considered hate speeches:

1 – *Reciprocal presupposition*. The two isotopies can be considered as one isotopy. The hate speeches are featured by a weak, classic correlation ( $0 < C < 0.5, 0 < S < 1.4$ ).

2 – *Dominance*. The presupposition is unidirectional (one of the two lexemes is incidental). The hate speeches are featured by no correlation or a weak, classic anticorrelation ( $-0.5 < C < 0, 0 < S < 1.4$ ).

3 – *Distinctiveness*. The two isotopies are well individuated, and they do not overlap in the considered hate speech. The hate speeches are featured by a strong, classic anticorrelation ( $-0.7 < C < 0.5, 1.4 < S < 2$ ).

4 – *Allotopy*. The two lexemes are allotopic: they simply do not share the same contexts. They are strongly opposed. The hate speeches are featured by a strong, quantum anticorrelation ( $-1 < C < 0.7, 2 < S < 2.8$ ).

Interestingly, we found a link between the correlation value and the bell value, so that only some strong anticorrelations violate the Bell inequality. This point will be discussed below.

#### 4.1 Reciprocal presupposition

The first discursive subset (fig. 4) corresponds to a weak classical correlation between the two isotopies. This corresponds to a general topic of the hate speech where the two lexemes represents two intersecting sets: for example, *black women*.

Correlation	Bell Value	Topic
Weak: $0 < C < (-0.5)$	Classic: $0 < S < 1.4$	Intersection of the two isotopies. Example: black women
Example: Based on the many, many videos I've watched of chimpouts, black women are more aggressive and more violent than black men. They seem to think there are no consequences for them when they punch other people in the face		

Fig. 4 - first discursive subset of the hate speeches: intersection between isotopies (in the example: black women)

The correlation value indicates the presence of a weak correlation between the two terms. They are not used as synonyms; rather, there is a presuppositions in terms of Greimas' square. For example, the considered hate speech oppose black men to black women, subdividing the presupposed black set in two presupposing subsets.

#### 4.2 Dominance

The second discursive subset (fig. 5) corresponds to the absence of correlation or to the presence of a weak anticorrelation between the two isotopies. The Bell value is still classical and weak ( $S < 1.4$ ). This corresponds to a general topic of the hate speech where the one of the two lexeme dominates the other, which is used incidentally.

Correlation	Bell Value	Topic
No correlation or weak anticorrelation $-0.5 < C < 0$	Classic: $0 < S < 1.4$	Dominance of one of the two isotopies. Example: women. Incidentally, white women
Example: Those 20 women ought to be quarantined in a special zoo and denied treatment for their HIV. Then every white woman should be forced to walk through that zoo to see those women slowly die from race-treason. These whorish women need to be brought back into line, they will be the death of our race.		

Fig. 5 - second discursive subset of the hate speeches: dominance (in the example: White and Women)

The correlation value indicates the presence of no correlation or of a weak anticorrelation between the two terms. They are not used as antonyms; rather, one of them prevails on the other. For example, the considered hate speech speaks about women. Incidentally, it makes references to white women.

### 4.3 Distinctiveness

The correlation value indicates the presence of a strong anticorrelation between the two terms (Fig. 6). The Bell Value is higher than in the previous subsets, but it is still classical ( $S < 2$ ). There is no intersections between the two isotopies: they are well individuated and kept distinct. For example, the considered hate speech accuses liberals of partisanship about women *and* about black people speaks about women.

Correlation	Bell Value	Topic
Strong anticorrelation $-0.7 < C < 0.5$	Classic: $1.4 < S < 2$	It rises from the distinctiveness of the two isotopies. Example: black people and women.
Example: Liberals only teach the bad in American history. I had multiple teachers that told me that slavery affects black people today and women only make 70 cents to a man. These are both lies, and there is nothing taught about how we spread ideas of individual freedom across the western world and gave more rights to women, minorities, plants and animals than any other, all thanks to “racist slave holders” so yeah, teach slavery all you want, but also include the fact that these ideas were not constitutional and mostly pushed by democrats.		

Fig. 6 - third discursive subset of the hate speeches: distinctiveness (in the example, between black people and women)

### 4.4 Allotopy

The last, very interesting case, is represented by the presence of the strongest anticorrelation and a quantum Bell Value

Correlation	Bell Value	Topic
Strong anticorrelation $-1 < C < (-0.7)$	Quantum: $2 < S < 2.8$	Is the result of the allotopic relation between the considered lexemes. Example: women Vs. hate
Example: >>>Glad you think a man raping a woman is an “equally likely scenario” as a woman drunkenly hitting on and having sex with a man.  Fuck off, and take your hate elsewhere. Edit: is this r/feminism now? Did no one read what this bitch wrote?  >>>But there is the equally likely scenario where the woman gets drunk, and a man steps in to “take care of her”. Separates her from her friends, says he’ll walk her home.  Obvious man hater here.		

Fig. 7 - fourth discursive subset of the hate speeches: (in the example: women vs. hate)

The example is very interesting too, since the writer ‘quotes’ the discourse of the interlocutor. The first focus on hate (‘take your hate elsewhere’, ‘man hater here’, while the second focus on women and men). The lexemes ‘women’ and ‘hate’ are allotopic: they do not share the same contexts.

### 5. Discussion

The particular link we found between correlation and bell value probably depends on the features of the textual genre we analyzed: hate speeches are indeed short, lexically poor, violently opposing two or three terms. Thus, further research is needed to fully understand whether the typology we individuated is complete and relevant to other textual genres. For example, strong positive correlations are not present in the corpus; this does not mean that they are not possible. Furthermore, a comparison between hate and non-hate speeches could lead to a better understanding of the difference between them.

a better understanding of the difference hate/non hate speeches;

An interesting point concerns Quantum anticorrelations, because they suggest that formal semantic models should be weaker than classical logic. In fact, comparison to ordinary logic, quantum logic is an extended system (Von Neumann 1932:253). For example, let us see another hate speech, opposing white (men) to (feminist) women and to Black (table 8):

Correlation	Bell Value	Topic
Strong anticorrelation -1 < C < (-0.7)	Quantum: 2 < S < 2.8	Women is opposed Black Black is opposed to White White is opposed to Women
Example: Sometimes I feel like those movements became obsolete the moment women got equal rights with men and people stopped thinking about blacks as of inferior race. Now they just keep momentum, turning women and minorities into privileged classes. If they keep this up in a few decades we would *need* MRA and white rights activists.		

Fig. 8 - In this text we find a strong anti-correlation between Women, Black, and White respectively. At the same time, all the considered relations violate Bell inequality

Our algorithm registers three strong quantum anti-correlations: white/men, white/women, women/black, which seems adequate to our interpretation of the message. However, this seems a violation of first order propositional logic. In fact, in classical logic, (a) is a tautology:

a)  $(A \leftrightarrow \neg B) \wedge (B \leftrightarrow \neg C) \rightarrow (A \leftrightarrow C)$

Let A = the ‘women’ isotopy, B = the ‘white’ isotopy, and C = the ‘Black’ isotopy. Thus, “if (Women iff not White) and (White iff not Black) then (Women iff White)” is always true. We could call this rule ‘the enemy of my enemy is my friend.’ In our case, this would imply that women and black would be somehow correlated isotopies: this does not happen in the considered texts, where the three lexemes are respectively allotopic and /women/ and /black/ produce non overlapping isotopies. The reason of the difference between quantum anti-correlation and classical logic consists in the geometry of the considered space. In Galofaro, Toffano, and Doan (2018) we demonstrated how anti-correlation is related to the angle between the base-vectors of

the query. Roughly speaking, “being anti-correlated” equals to “being orthogonal”, and “being correlated” equals to “being parallel”. Let us interpret (a) in geometrical terms:

b) If “Women” is orthogonal to “White” and “White” is orthogonal to “Black”, then “Black” is parallel (and, a fortiori, not orthogonal) to “Women”;

The sentence (b) would be true in a two-dimensional semantic space. In our space, each vector of the document (white, women, black ...) lays in a different dimension, since they are all orthogonal. Thus, if all the three base-vector are anti-correlated, we can represent them as in Fig. 9.

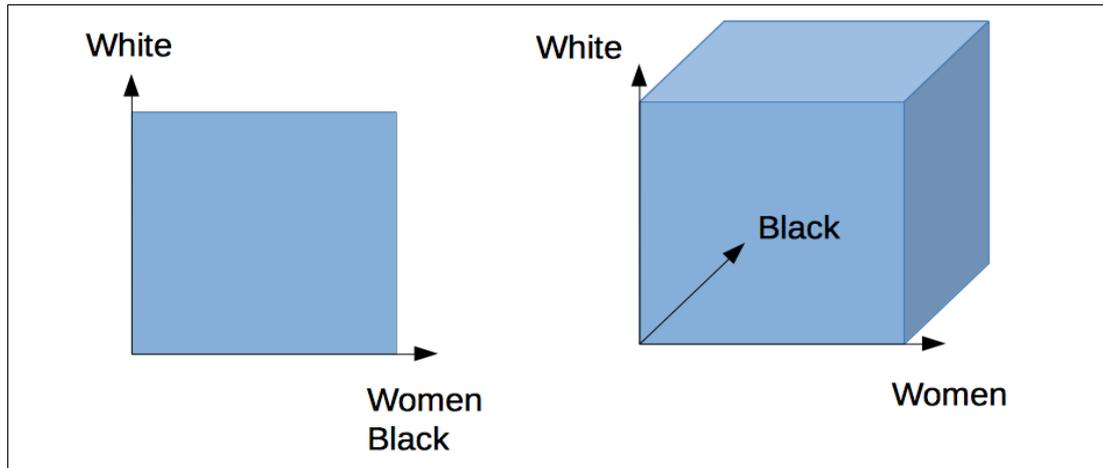


Fig. 9 - A geometrical interpretation of anti-correlation in a 2D and in a 3D space

## 6. Open questions

Why semantic space should be represented as the same space of quantum computation and quantum physics? Jean Petitot formulated the problem in this way:

Many people are using quantum formalisms beyond physics but it is in general difficult to justify the Hilbert structure (in particular complex coefficients with phase factors necessary for interferences) (Petitot, personal communication, 2018).

The first answer could have been: whatever works. Or: mathematics is just a formal model, the fact that a portion of semantics and physics can be formalized using the same tools does not suggest any ontological relation between the two. To paraphrase non-realist interpretations of quantum logic – see Wilce (2017) – quantum semantics is a theory about the possible statistical distributions of lexemes in certain contexts, and its non-classical “logic” simply reflects the fact that these distributions can not be present simultaneously anywhere in the text. Because of this, the set of propositions on isotopies is less rich than it would be in classical probability theory, and the set of possible statistical distributions, accordingly, less tightly constrained, allowing cases as the one we reported in the previous paragraph. That some “non-classical” probability distributions allowed by this theory are actually manifested in nature is perhaps surprising, but in no way requires any revision of the semantics of the language we use to make reference to nature.

## 6.1 Semantics and Quantum information

However, there is another possible explanation. The point is the kernel notion of information: in particular, Von Neumann entropy. Just as Shannon entropy measures the amount of order in a classical system, von Neumann entropy measures order in a given quantum system. Von Neumann information is calculated by calculating the eigenvalues of a density matrix where we store the  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  components of the document vector  $|\Psi\rangle$  expressed in the two different bases provided by the two lexemes we are interested in (Yanowski and Iannucci 2008 : 288-295).

According to this point, when we measure the expectation that two lexemes are related in the same contexts, we are not ‘understanding’ the text. The same reason led Umberto Eco (1962) to understand that Information Theory did not provide a complete foundation for aesthetics, and to start his research on semiotics. We could do the same operation with an undeciphered writing, such as a linear A tablet. We are only acquiring information *on* semantics, i.e. on the relation between certain words. For example, this way we can understand that ‘schtroumpf’ is the opposite of a ‘schtroumpfette’, without actually knowing what a ‘schtroumpf’ is. However, this would be very helpful to decode an encrypted message. Further researches on the density operator are needed to a better understanding of the problem formulated by Petitot.

## 7. Conclusion

Quantum semantics tries to merge a structural notion of value as *difference* with a phenomenological notion of value resulting from the intentional relation between subject and object (inherence, cf. Marsciani 2014). In fact, the model foresees two levels (Fig. 9):

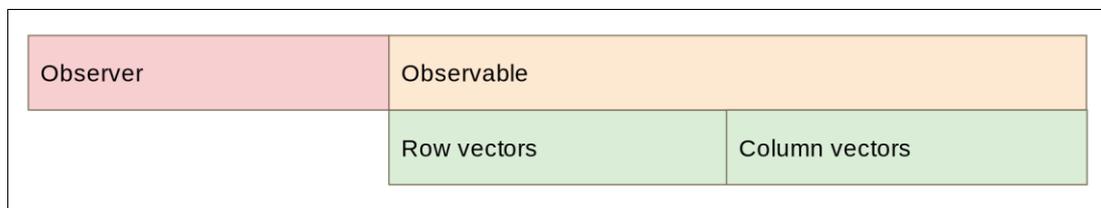


Fig. 9 - A two-level model

According to the model, meaning is produced by the relations between world-vectors in the semantic space of the document, but only in so far as it is observable. The observer interacts with the document transforming it and progressively determining it: meaning is transformation.

## References

Aa.Vv. (2012), «Countering hate speech online», in *EEANews*, from: <https://eeagrants.org/news/countering-hate-speech-online> last accessed: October 1<sup>st</sup> 2019.

Barros, Joao, Toffano, Zeno, Meguebli, Youssef, and Bich-Liên Doan (2014), «Contextual Query Using Bell Tests», in *Lecture Notes in Computer Science*, Vol. 8369, Springer, Berlin, pp. 110-121.

Dubois François, Toffano Zeno (2017) «Eigenlogic: A Quantum View for Multiple-Valued and Fuzzy Systems», in *Quantum Interaction. QI 2016. Lecture Notes in Computer Science*, Vol. 10106, Springer, Berlin, pp. 239-251.

Eco, Umberto (1962) *Opera aperta*, Bompiani, Milano 1976 (*Open work*, transl. by A. Cancogni, Harvard University Press, Cambridge Mass. 1989)

Eco, Umberto (1972), *Le forme del contenuto*, Bompiani, Milano.

Ferraro, Guido (2019) *Semiotica 3.0: 50 idee chiave per un rilancio della scienza della significazione*, I saggi di Lexia 31, Roma, Aracne.

Galofaro, Francesco, Toffano, Zeno and Bich-Liên Doan (2018), «Quantum Semantic Correlations in Hate and Non-Hate Speeches», in *Electronic Proceedings in Theoretical Computer Science*, Vol. 283, pp. 62-74, from <http://eptcs.web.cse.unsw.edu.au/content.cgi?CAPNS2018>

Greimas, Algirdas Algirdas J. (1966), *Sémantique structurale*, Paris, PUF 2002 (*Structural semantics: an attempt at a method*, transl. by D. McDowell, R. Schleifer, A. Velie, University of Nebraska Press, Lincoln NE 1984).

Greimas, Algirdas J., Rastier, François (1968), «The interaction of semiotic constraints», in *Yale French Studies*, 41, pp. 86-105.

Lund, Kevin and Burgess, Curt (1996), «Producing high-dimensional semantic spaces from lexical co-occurrence», in *Behavior Research Methods, Instruments, & Computers*, Vol. 28, n. 2, pp. 203- 208.

Marsciani, Francesco (2014), «À propos de quelques questions inactuelles en théorie de la signification», in *Actes sémiotiques*, n. 117, from <https://www.unilim.fr/actes-semiotiques/5279>

Nockleby, John T. (2000), *Hate Speech*, in Levy, Leonard W., Karst, Kenneth L. *et al.*, *Encyclopedia of the American Constitution*, Macmillan, New York, pp. 1277-1279.

Quillian, M. Ross (1968), *Semantic Memory*, in Minsky, Marvin (1968), *Semantic Information Processing*, Cambridge Mass, MIT press.

Rastier, François (2009), *Sémantique interprétative*, Presses Universitaires de France. 3<sup>rd</sup> edition.

Van Rijsbergen, C. J. Keith (2004), *The Geometry of Information Retrieval*, Cambridge University Press, Cambridge.

Von Neumann, 1932, *Mathematische Grundlagen der Quantenmechanik*, Springer-Verlag, Berlin (*Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton 1955).

Wilce, Alexander (2017), *Quantum Logic and Probability Theory*, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/spr2017/entries/qt-quantlog/>

Yanowski, Noson S. and Iannucci, Mirco A. (2008), *Quantum Computing for Computer Scientists*, Cambridge University Press, Cambridge.

Ziqi Zhang & Lei Luo (2018), «Hate Speech Detection: A Solved Problem? The Challenging Case of Long Tail on Twitter», in *Semantic Web*, in press (status: accepted) from arXiv preprint arXiv:1803.03662.