Quantum Semantic Correlations in Hate and Non-Hate Speeches

Francesco Galofaro, Zeno Toffano, Bich-Liên Doan

Quantum Semantic Correlations in Hate and Non-Hate Speeches*

Francesco Galofaro[1,2], Zeno Toffano[3], and Bich-Liên Doan[3]

[1] Politecnico di Milano, 20158 Milano, Italy
[2] Free University of Bozen, 39100 Bozen, Italy
francesco.galofaro@polimi.it
[3] CentraleSupélec, 91190, Gif-sur-Yvette, France
{zeno.toffano,Bich-lien.doan}@centralesupelec.fr

Abstract. This paper aims to apply the notions of quantum geometry and correlation to the typification of semantic relations between couples of keywords in different documents. In particular we analysed texts classified as hate / non hate speeches, containing the keywords *women*, *white*, and *black*. The paper compares this approach to cosine similarity a classical methodology, to cast light on the notion of "similar meaning".

Keywords: Quantum Information Retrieval · Semiotics · Digital Humanities.

1 Corpus

The Online Hate Index [1], a joint initiative of Anti Defamation League's Center for Technology and Society and UC Berkeley's D-Lab, consists of 7619 comments from the platform Reddit, collected in 2016 during the USA Presidential campaign. While the corpus has been funded by ADL, no research activities are conducted by ADL or directed based on their personal goals: research design and testing has been conducted entirely by Berkeley D-Lab, an interdisciplinary academic unit. The comments were labelled by a team of undergraduates under the supervision and using methods developed by D-Lab. ADL had no involvement in the selection of comments, development of methods, and labels applied to comments. 7184 comments have been manually labelled "non-hate", whereas 411 have been considered as "hate" speeches with the purpose of stimulating further machine learning researches. The five most used words in the collected hate speeches are *Jews*, *white*, *hate*, *black*, and *women*. Hate speeches present also some peculiar features: the average number of words, the average number of all caps, and the average sentence length is higher. Among the five most used words in hate speeches, *white* and *black* are interesting because they can be considered an antonymic couple from the point of view of lexical semantics. For this reason, we decided to analyse the semantic relations between the terms *white*, *black*, and *women*.

* This paper has been made possible thanks to a corpus of hate and non-hate speeches shared by the Berkeley D-Lab.

1.1 Sub-corpora

As the corpus is subdivided into hate (see table 1) and non-hate speeches (see table 2), we identified six sub-corpora:

- H.WBWh: hate speeches corresponding to the logical expression $black * white * women$;
- H.WhW-B: $white * women - black$ hate speeches;
- H.BW-Wh: $black * women - white$ hate speeches;
- NH.WBWh: $black * white * women$ non-hate speeches;
- NH.WhW-B: $white * women - black$ non-hate speeches;
- NH.BW-Wh: $black * women - white$ non-hate speeches;

As we can see, selected hate-speeches tend to use the words *white*, *black*, and *women* in the same document, whereas not-hate speeches tend to use only *white*, *women* or *black*, *women*. It is sometimes difficult to understand the criteria that have been used for labelling the corpus. In particular, NH.WBWh.2 and NH.WhW-B.4, two documents labelled as non-hate, though avoiding offensive terms contain stereotypes on the relation between race and IQ. We hope that a semantic analysis will be helpful to re-discuss the classification in case of mistakes, e.g. if a document classified as non-hate speech shows the same semantic features of hate speeches.

1.2 Note on the formalism

The logical formalism we used to represent elementary semantic relations corresponds both to Boole's algebra and to the one proposed by Greimas and Rastier in their well-known semiotic square [9] and suggests a strong connection to Dubois and Toffano's eigenlogic which could provide a formal description of the semantic relations immanent to the text, thus allowing the formulation of semantic queries [6].

Table 1: Hate speeches. Italic underlines the occurences of the keywords.

Query: $black * white * women$
H.WBWh.1 - Did you ever notice how these anti-*white* racist *black*s always pick on the old, the weak or the infirm? Old ladies. Children. Young *women*. Teenage geeks. And they almost never do it alone, but run in packs, like jackals or hyenas.
H.WBWh.2 - Eastern Europe will likely be the most reliable backup plan for most *white*s. And while most of western Europe is cucked out right now, I have higher long term hopes for there than America. Most Americans are satisfied with only going so far as civic nationalism. Nationalism is intertwined with ethnicity in Europe and so ethnonationalism is the inevitable result from the resurgence of the right there. I expect that America's nationalism will never again become attached to race, unfortunately. The elites in all western countries are banking on a multiracial future and in many cases are mating accordingly. This girl has mostly *white* features and in all likelihood is 75% *white* (given that most American *black*s are 20-40% *white* themselves), but the media will play up her *black*ness in order to make it seem that *black women* are desirable.

H.WBWh.3 - It was always a lot like this. There were always elements of it. For example it was immidiately anti-male from the word go. It was just easier to campaign that way because society was so anti-male / pro-female already. Hence the saying that feminism is just the other side of the coin from traditional conservatism. For example feminists went on and on about drunk husbands beating their poor wives the way they go on about rape today. A good false accusation makes for fine propaganda and feeds into what everyone likes to think about men anyway. Feminists could have taken the high road. They didn't. They took the easy screw men over, sex war men and *women* are enemies and men are all evil road. So the rot was there from the beginning. So was the narcissism. The victimry. The "we rich *white women* are worse off than *black* slaves" crap. Reality was they were treated better than anyone else. but that doesn't sell well. Victimry does (for *women*). They shamelessly used their power to get more. Sure some wanted equal responsibilities and duties too at first but those voices, never a majority, got fewer and fewer. Getting more privileges was just the easy path. So yes to the vote, but no to conscription that went with it. And you can see how close to a hate movement all that already was even back in the day when they had some issues that were actually worth fixing (even if they were not a matter of equality because men's issues were not being addressed too). But since they got all their issues dealt with by around 1890 (except the vote because most *women* still opposed the vote for *women* until the 1910s) they became a movement without a cause and just wnet off the deep end of anti-male hate.
H.WBWh.4 - Sometimes I feel like those movements became obsolete the moment *women* got equal rights with men and people stopped thinking about *black*s 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.
Query: $white * women - black$
H.WhW-B.1 - 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.
H.WhW-B.2 - Jewish gangs kidnapping/raping/killing *white* children as young as 2 years old to sell snuff films](http://www.jewwatch.com/jew-atrocities-blood-libel.html) , then Jews had the Italian news executives fired for AIRING THE STORY (Jews HATE being called out on crime and DEMAND FULL CONCEALMENT or else it's "ANTI SEMITIC!!!"....Jews think it's ANTI SEMITIC to PROTECT YOUR CHILDREN FROM THEM). [Israel's $1 billion a year sex slave trade of *white women*](http://www.jewlicious.com/2006/04/because-we-were-slaves-israels-sex-trade/)
H.WhW-B.3 There is no need to side with kikes against *white women* who put their money where their mouth is and goes and protests the jew at her own risk. Why on earth do I care if the Jews are weakening someone who is an avowed enemy, like Hamas? Granted they're not capable of it, but they've made declarations of intent for conquering Italy and other places in Europe. The Muslims will forever wage their jihad against Europe regardless of whether or not Israel exists. Their religion is one of rape and plunder. I don't really protest Israel because I couldn't care less what Israel does in Israel. If the time comes for Europe to reclaim Turkey and the Levant, then I will care. I care what Israeli stooges do here in the west, what propaganda they sow, what legislation they lobby for, etc.
Query: $black * women - white$
H.BW-Wh.1 - 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.
H.BW-Wh.2 - 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.

Table 2: Non-hate speeches. Italic underlines the occurrences of the keywords.

Query: $black * white * women$
NH.WBWh.1 - I think mansplaining might be a real thing. Explaining done by a man is a real thing. But why make it gendered? *women* explain things to men all the time, at times also using a patronizing tone. Anecdotally, I've experience more *women* explaining something in a patronizing way than I have men. Even if there would be statistics that show men do it more often in a patronizing way (which there aren't), it's still hard to argue for making it gendered. To put it in an analogy: we know that *black* people in the USA commit more burglary than *white* people (in relative terms). Should we call it "*black*burgling"? I don't think so: I think it implies that comitting burglary is somewhat characteristic of *black* people, which isn't the case since it's only a minority of *black* people doing it. It would probably also help drive the wedge even further between *black* and *white* people. And what would have been gained? A fancy new word to insult *black* people with? Why can't it just be "burglary (done by a *black* person)"? The same goes here.

NH.WBWh.2 - Well, you're not wrong. *black*s, men and *women*, are of significantly lower intelligence than all other races on the planet, with the single exception of the aborigines of Australia, who are just as limited, mentally, as *black* Africans. I'm talking about a BIG difference in intelligence, not a small difference. The average IQ of *white*s is around 100. Hispanics come in at around 89. American *black*s, who have mingled their genetics with *white*s for generations of interbreeding, come in at an average of 85. African *black*s have an average IQ of 70! That's right, 70. By normal *white* standards, the average negro in Africa is mentally retarded. These IQ tests have been done many different times, many different ways, all around the globe, and they all show the same thing. They are not wrong. Since you are *black* yourself, I should point out that a general or average IQ for a race has no bearing on the IQ of any individual in that race. By that I mean you yourself may be a genius. There are many *black* geniuses. I am talking here about average intelligence. Why is average intelligence important? Because it dictates what a race, as a whole, is able to accomplish. If you look at Africa, *black*s were able to build almost nothing. No roads. No wagons to drive on those non-roads. They didn't even learn to domesticate horses to ride by themselves. Everything African negroes have, they learned from other races. And please don't talk about Egypt. The peoples of northern Africa were not negroes in ancient times, and to a large extent are still not negroes. They were Mediterranian peoples, like the Greeks, Etruscans and Romans. Why is the lack of accomplishment of African *black*s in history important today? Because *black*s still cannot create or build anything of importance, as a race. Their low intelligence, coupled with other negative factors that have been less well-demonstrated, prevent them from achieving anything. Just look at any *black* city in America. Any *black* city, take your pick. As soon as a city goes *black*, it goes to ruin and decay. Always.

Query: $white * women - black$

NH.WhW-B.1 - 8chan trolls are using it to abuse *women*, children, minorities, and gays Hetero *white* men can't be abused, folks. If the very same thing happens to them it's... well... irrelevant.
NH.WhW-B.2 - Many *white women* voted for Trump not because they were concerned about their income, but because they are the ?Ivanka voters?. Their vote was for Ivanka instead of Donald. They love her style and success story. I can't imagine this being true at all.
NH.WhW-B.3 - Then there's *white* ribbon day. With posters saying 'Are you man enough to stop violence against *women*?' http://*white*ribbon.org.nz/
NH.WhW-B.4 - It's become a sort of cliche that Hispanic *women* apparently urge their daughters to seek out *white* men (or at the very least, lighter skinned men) in order to "improve the race". There is about a 10 point IQ difference between *white*s and most mestizos, and fairer skin is usually seen as more attractive regardless of political beliefs, so their efforts aren't misguided. The last time I was in Corpus Christi, as well as when I visited Phoenix and San Diego, it seemed like every single young *white* male had a mestizo girlfriend. And in their minds "why not?" An average *white* guy can get a high end mestizo girl easier than an average *white* girl, and chances are that she is eager to please because she is happy just to have a *white* guy. The rate of WM/HF mixing in the southwest is every bit as bad as WM/AF mixing on the West Coast. Here in the Midwest the most commonly seen mixed couple is BM/WF so I was rather shocked when I've gone out west and seen the extent of what is happening out there.
NH.WhW-B.5 - How would energising the base help. He already has them. "You would be in jail" "hate in your heart" and the "devil" comment are all for the base and he still dropped in the polls afterwards. The guy needs greater support from *women*, he needs greater support from minorities. As far as I know registration hasnt surged for the non college educated *white* voters so he needs that as well. Currently clinton can run out the clock, she won't and isn't trying that but she could rather comfortably not do much and let trump focus on attacking his accusers. I have a feeling he is going to leave that arguement behind now given that he has called the accusers horrible horrible liars and ?Take a look. Look at her. Look at her words. And you tell me what you think. I don?t think so,?
NH.WhW-B.6 - I wonder what she would say to me. I'm a *white* male that voted for Jill Stein/Ajamu Baraka. I voted for 2 *women*, rather than her 1.

Query: $black * women - white$

NH.BW-Wh.1 - Win or lose, a lot of people have got the red pill. Trump just did not play the demographics well. Alienating *black*s, mexicans, muslims, *women*. But many of them also know what is happening, but they consider trump worse The establishment has suffered from a huge loss in credibility which they can never recover. As an example see the comments on cnn, abc, nyt facebook pages
NH.BW-Wh.2 - Unlike jubbergun, i didn't even notice the race of the attackers. I think one of the men standing around was *black*, but that was the only time i noticed color. Jubbergun is either a troll or a PC moron, that's all i can say. As for my specifically using bonobos rather than common chimps or gorillas, that is because the bonobos are our closest relatives and also one of the few species of primate amongst which the females commit most of the violence. Hence, these *women* have unleashed their inner bonobo.

NH.BW-Wh.3 - I'll admit, I'm really pro sanders but this upset me. He was pretty rude to the woman, I wish he let her talk more and had more of a conversation. To me, most the arguments I get into, the one who rages first is often the one with at least the most to lose and at most the least willed debater. Not to mention it is prejudice. Urban? Come on. If humans fight, if we're in a war, we use guns. Plain and simple. The *black* men in this country and *black women* in particular have been SHUT OUT of so many opportunities in this country and MANIPULATED into the war they think they have to wage. Manipulated by terrible, evil motherfuckers in badges who go around thinking they're doing the lord's work. And really, really fucking unfortunately, I can't help but feel that Sanders' view as the socialist democrat is so far right and so agreed upon by so many... it really does show how ingrained gun culture is in our country.
NH.BW-Wh.4 - claims quoting Dr. Martin Luther King, Jr., to *black women* is a violent and ?cisheteropatriarchy? act. You seriously can't make this up.
NH.BW-Wh.5 - That's probably because 30 years ago they were not bashing *black*s or *women*. Well, *women* only got bashed if they mouthed off.

2 State-of-the-art

A geometric approach to semantic space studies has been proposed first by Jean Petitot, in terms of catastrophe theory [14]. As quantum geometry is concerned, it is used by scholars in Information Retrieval for the purpose of unifying vector, logic, and statistical approaches [19, 12]. Among others, Bruza and Woods [3], applied it to the semantic representation of polysemic words as a superposition state. Barros, Toffano, Meguebli and Doan [2] proposed to interpret the notion of entanglement as a measure of the strength of the semantic relation between two query-words, both present in a certain document. To this purpose, using the Hyperspace Analogue to Language (HAL) method [11], the authors formalised the semantic space of a document as a square matrix, as we will explain hereafter. Many quantum information retrieval scholars prefer this technique because it is Hermitian and it allows the implementation of a density matrix [19, 12]. Instead of measuring cosine similarity between two keywords, the work in [2] makes use of the Gram-Schmidt orthogonalisation method to measure the degree of correlation between the words, characterized by the violation of a CSHS inequality [4]. Pushing forward this idea, Galofaro, Toffano, and Doan [8] proposed a theoretical paper in which observables are interpreted as semantic features. The Born rule is used to find the expectation values associated to the application of a specific observable to two word-vectors in order to measure the degree of correlation/anticorrelation between them [18]. The present paper aims to test this method, and to compare it with the classical cosine similarity measure.

3 Relevance to language processing

It is possible to ask how are we going to interpret the correlation value in terms of linguistic features. According to Umberto Eco [7], the terms "semantics" has been used in five different acceptations:

1. Lexicology: a study of meaning outside every context (dictionary);
2. Structural Semantics: interested in semantic fields considered as systems;
3. Study of the relation between the meaning and the referent;
4. Truth-conditional logic;

5. Textual semantics: a study of the peculiar meaning assumed by terms and words in their context;

Though the five levels are obviously related, the text and the context have the last word in defining the meaning of terms. For example, according to any thesaurus, *black* and *white* are antonyms (if black, then not white and vice versa). Having a look at our corpus, we find: "most American *blacks are* 20-40% *white*" (H.WBWh.2), weakening the antonymy. The HAL method allows us to work on semantics in sense of 5 because we formalise the semantic relations between terms that constitute a given context. These become the characteristics of a semantic space.

3.1 Commutation test and quantum correlation

With measuring the degree of quantum correlation we are searching for a semantic equivalent to Hjelmslev' commutation test. Commutation of elements of the *expression plan* aims to search for linguistic units. If we substitute "black" with "Afro-American" in *blacks, men and women are of significantly lower intelligence than all other races on the planet*, we notice how meaning is unaltered, while if we substitute it with "Afghan hound", the meaning changes. We could even suggest that this is why the original sentence is actually racist: we speak about men as they were dogs. However, **Afghan hounds, men and women* is not correct in English because of structural reasons related to semantics in sense 2: "men" and "women" carry a structural *classeme* (an element of meaning) $(human \rightarrow -animal)$ [10] . What if we were able to commute *meanings*, and not *signifiers*? For example, what if we were able to change the meaning unit "human" in "dog" while preserving "male" and "female" all along the sentence? This is what we mean by "commutation test on the content plan". As a result of the test, such an abstract machine as a computer could probably generate, on the expression plan, a sentence like *Afghan hounds, sires and bitches*. The Born rule provides a tool to measure the expectation values for these commutations.

4 Design

In synthesis, we prepared the corpus reducing each word to its stem; we then applied the HAL method to obtain two word vectors representing the keywords we are interested in, and a document vector; finally, we measured the the cosine similarity of the keywords and their (anti)correlation value.

4.1 Cleaning the corpus

Since we are interested in every kind of semantic information not manifested by morphology or syntax, we used the Python library *nltk Lancaster* stem to reduce different tokens to the same type (e.g. black, blacks, blackness). The *Lancaster* stemmer is more aggressive than the alternative *nltk Porter* stemmer, which can

distinguish between *woman* and *women*. Obviously, a stem is not necessarily identical to its morphological root: our purpose is only to reconstruct the immanent net of relations underlying the manifest words. For a similar reason, we used ntlk stopwords list to eliminate syncategorematic terms. We also used regex to eliminate all not relevant signs such as punctuation [20].

4.2 The matrix

We applied the HAL method to each document of each sub-corpus to formalise it. Given k roots occurring in the document, we calculate a $k \times k$ matrix which represents the semantic space of the document.

Example 1. Table 3 shows the HAL matrix of the document:

NH.BW-Wh.5: *That's probably because 30 years ago they were not bashing blacks or women. Well, women only got bashed if they mouthed off*

Table 3. HAL Matrix corresponding to document NH.BW-Wh.5 (window: 11)

	30	**ago**	**bash**	**becaus**	**black**	**got**	**if**	**mouth**	**not**	**off**	**onli**	**or**	**probabl**	**s**	**that**	**they**	**well**	**were**	**women**	**year**
30	10	0	0	9	0	0	0	0	0	0	0	0	8	7	6	0	0	0	0	0
ago	8	10	0	7	0	0	0	0	0	0	0	0	6	5	4	0	0	0	0	9
bash	4	6	22	3	3	9	0	0	10	0	8	4	2	1	0	7	6	8	12	5
becaus	0	0	0	10	0	0	0	0	0	0	0	0	9	8	7	0	0	0	0	0
black	3	5	9	2	10	0	0	0	8	0	0	0	1	0	0	6	0	7	0	4
got	0	0	3	0	4	10	0	0	2	0	9	5	0	0	0	0	7	1	14	0
if	0	0	10	0	2	8	10	0	0	0	7	3	0	0	0	0	5	0	10	0
mouth	0	0	7	0	0	6	8	10	0	0	5	1	0	0	0	9	3	0	6	0
not	5	7	0	4	0	0	0	0	10	0	0	0	3	2	1	8	0	9	0	6
off	0	0	6	0	0	5	7	9	0	10	4	0	0	0	0	8	2	0	4	0
onli	0	0	4	0	5	0	0	0	3	0	10	6	0	0	0	1	8	2	16	0
or	2	4	8	1	9	0	0	0	7	0	0	10	0	0	0	5	0	6	0	3
probabl	0	0	0	0	0	0	0	0	0	0	0	0	10	9	8	0	0	0	0	0
s	0	0	0	0	0	0	0	0	0	0	0	0	0	10	9	0	0	0	0	0
that	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
they	7	9	8	6	1	7	9	0	0	0	6	2	5	4	3	20	4	0	8	8
well	0	2	6	0	7	0	0	0	5	0	0	8	0	0	0	3	10	4	9	1
were	6	8	0	5	0	0	0	0	0	0	0	0	4	3	2	9	0	10	0	7
women	1	4	12	0	14	0	0	0	10	0	0	16	0	0	0	6	9	8	28	2
year	9	0	0	8	0	0	0	0	0	0	0	0	7	6	5	0	0	0	0	10

Each square matrix is calculated moving a window, representing the considered context, over the document, stem by stem. All stems within the window co-occur with the last stem with a strength which is inversely proportional to the distance between the stems. We finally sum the different occurrences of stems: for example, women occurs two times in our document.

4.3 Cosine similarity

In a HAL matrix, rows and columns differ. A word-vector is then represented by the concatenation between the corresponding row and column vectors. In this

way we obtain the word-vectors of the keywords we are interested in (white, black, women): $|w_{white}\rangle$, $|w_{black}\rangle$, $|w_{women}\rangle$. We can now calculate the angle between any two word-vectors as well as their cosine similarity (cs), since "cosine has the nice property that it is 1.0 for identical vectors and 0.0 for orthogonal vectors" [16]. Usually, cosine similarity measures the similarity between the query vector and the document vector. For this reason, the way we use it, measuring cosine similarity between the keywords in each document, and obtaining each time a different measure, could seem rather unorthodox. However, since the two keywords are just vectors, their angle can be used to measure their similarity in the particular semantic space corresponding to a certain document. For example, as Song, Bruza, and Cole wrote, "nurse and doctor are similar in semantics to each other, as they always experience the same contexts, i.e., hospital, patients, etc." [17]. The reason why we choose to compare cosine similarity to the expectation degree measured through the Born rule is perhaps not intuitive. In both case we deal with many-dimensional vectors, and not with punctiform events. For this reason we will not consider the euclidean distance to calculate similarity or different methods to calculate frequency, such as pointwise mutual information (PPMI).

4.4 Gram-Schmidt orthogonalisation

In order to measure correlation between two keyword-vectors, let us say *black* and *women*, we first obtain a document vector $|\Psi\rangle$ summing all word-vectors. The next step is to apply the Gram-Schmidt orthogonalisation process to $|w_{black}\rangle$, $|w_{women}\rangle$ in order to obtain two pairs of orthonormal bases $\{|u_{black}\rangle, |u_{black\perp}\rangle\}$ and $\{|u_{women}\rangle, |u_{women\perp}\rangle\}$. If we project and normalise the document-vector $|\Psi\rangle$ onto each couple of bases we obtain a vector $|\phi\rangle$:

$$|\phi\rangle = \alpha\,|u_{black}\rangle + \alpha_{\perp}\,|u_{black\perp}\rangle = \beta\,|u_{women}\rangle + \beta_{\perp}\,|u_{women\perp}\rangle \tag{1}$$

We want to emphasize that we represented the document vector through its components on the two bases provided by each keyword-vector.

4.5 Abstract machines

The notion of abstract machine links quantum theory [18] to post-structuralist perspectives on meaning [5]. To typify the semantic relation between *black* and *women* in our corpus, we design two abstract machines: σ and τ, two linear operators. Their input vector is $|\phi\rangle$ (representing the document. The σ-machine operates on each context, and returns the output $+1$ when it acts on the vector $|u_{black}\rangle$ (representing the meaning of the stem *black*), -1 in the other case. In a similar way, the τ-machine applies the same transformation on the meaning of the stem *women*. Now let us imagine what happens when we apply both the machines to the document: $\sigma\tau\,|\phi\rangle$. Principally, we deal with three situations:

1. the two outcomes are correlated in every context, when the first output is $+1$ and the second is $+1$, and when the first is -1, the second is -1. If we multiply the two numbers we always score $+1$;

2. the two outcomes are anti-correlated: in every context where the first output is $+1$, the second will be -1 and also the other way round. If we multiply the two numbers, we will always score -1;
3. the two outcomes are not correlated. in some contexts the output of the two machines will be $\{+1,+1\}$, while in others it will be $\{+1,-1\}$, $\{-1,+1\}$, or $\{-1,-1\}$. The average of the outcomes in the different contexts of the considered document will be 0;

The three considered cases are extreme situations: we will also find weak correlations, in which the score will tend to 1, weak anti-correlation, where the score will tend to -1, and also absence of correlation, giving results near 0. The outcome of a generic machine can be a transformation or not, $+$ or $-$. Since we have two machines, we deal with a four-state semantic space $\sigma\tau = \{++,+-,-+,--\}$. To construct an example of a general machine, we use the following Pauli spin matrix:

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2}$$

The effect of this matrix is to switch the components of the state-vector to which it is applied. It is the equivalent of the logical gate *negation* in Quantum Computation [13]. In this way we define an operator $\hat{B}_x$ in the *black-base* $\{|u_{black}\rangle, |u_{black\perp}\rangle\}$, and an operator $\hat{W}_x$ in the *women-base* $\{|u_{women}\rangle, |u_{women\perp}\rangle\}$. As a result, for example, $\hat{W}_x$ switches all the $|u_{women}\rangle$-related components of $|\phi\rangle$:

$$\hat{W}_x(\beta\,|u_{women}\rangle + \beta_\perp\,|u_{women\perp}\rangle) = \beta\perp\,|u_{women}\rangle + \beta\,|u_{women\perp}\rangle \tag{3}$$

$\hat{B}_x$ acts in the same manner on the $|u_{black}\rangle$-related values of $|\phi\rangle$. To calculate the expected score of the application of both machines to the document-vector, we apply the Born rule, whose output will be a number between -1 and 1.

$$r = \langle\phi|\,\hat{B}_x\hat{W}_x\,|\phi\rangle \tag{4}$$

The more the two layers of meaning are connected, the more the two independent abstract machines will return similar outputs: thus we interpret r as the immanent correlation between the respective meanings expressed by the *black* and the *women* stems.

5 Comparing data

We measured cosine similarity and correlation in our corpus of hate and non-hate speeches corresponding to the logical query *black*white*women*. We measured similarity and correlation between *black*, *women* and *white*, *women* respectively. The window length varies from 4 to 10. The results are displayed in fig. 1, referring to hate speeches, and fig. 2 referring to non-hate speeches. Then we calculated the graphs corresponding to cosine similarity and correlation between *white*, *women* in the corpora of both hate and non-hate speeches where the term

black is absent - see figures 3 and 4. Finally, we applied the same procedure to *black*, *women* in the corpora of both hate and non-hate speeches where the term *white* is absent - see figures 5 and 6. To compare the results we focus on window lengths of 8-10, which seem associated to more stable values.

black* ∗ *white* ∗ *women Both hate and non-hate speeches present strong anticorrelation *black vs. women* and *white vs. women*. Looking at the document H.BWhW.4, it is not surprising to see that both *black/women* and *white/women* are anticorrelated, since the text draws a comparison between *women's rights* and *black's rights*. We notice also the correspondence between a $r \simeq 0$ correlation score and a $cs \simeq 0.7$ value of similarity: two word-vectors can be geometrically close without being correlated. Another interesting problem is the *black/white* opposition. Generally speaking, their anticorrelation is weaker than for the other two, and it tends to disappear in H.WBWh.1 Provided that lexical semantics would describe them as antonyms, it could seem strange that they are not anticorrelated in H.WBWh.1. According to textual semantics [10] semantic relations are modulated and transformed by their co-occurrence in contexts. In our document, *anti-, white,* and *racist* give a fundamental contribution to establish the contextual part of the meaning of *black*, providing it of contextual *classemic values* - Rastier calls them *afferent semes*, and distinguishes them from the *inherent semes*, which characterize the semantic nucleus of a *lexeme* [15]. In a similar way, H.WBWh.3 is also interesting, since it shows how two terms can be weakly similar ($0.5 \leq cs \leq 0.7$) and still weakly anticorrelated ($-0.5 \leq r \leq 0$).

white* ∗ *women* − *black Five out of six non-hate speeches present a strong anticorrelation *white vs. women*, whereas hate speeches are featured by a weaker anticorrelation; in one case, NH.WhW-B.2, we have a positive correlation, since the document focuses on *white women* voting for D. Trump. If we look at the other documents, they oppose *women* to white men, voters. We must point out how NH.WhW-B.4 could be considered a hate speech from a semantic point of view. In this text, *hispanic women* are opposed to *white girl, men, guys* and this explains the strong anticorrelation.

black* ∗ *women* − *white H.WB-Wh.1 presents a positive *correlation* between *black* and *women*: in fact the document opposes black women to black men without reference to white women (*women* → *black*). On the contrary, H.WB-Wh.2 present a strong anticorrelation *black vs. women*, since women and blacks are considered as two distinct minorities. Most non-hate speeches present a weak anticorrelation or a weak correlation, except for NH.WB-Wh.2, in which a maximal anticorrelation value is justified because the document is composed of two different sections, the first about black color and the second about women. In NH.BW-Wh.5 we can see again how a high score of similarity does not necessarily correspond to a correlation of sa given type: in this text, we have a first close co-occurrence of *women* and *black*; the second occurrence of *women* is free and it weakens the value of the first relation.

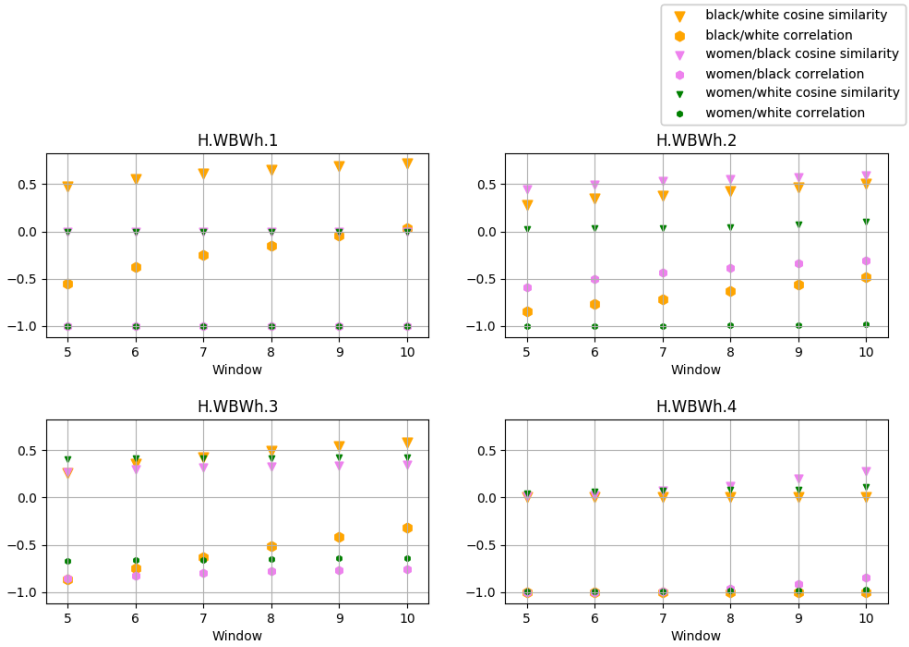

Fig. 1. Hate speeches H.WBWh.1-4. In H.WBWh.1, *black* and *white* show a high similarity score though they are not correlated. In the text, the meaning of "white" is modified both by the prefix *anti-* and by the presence of *black*.

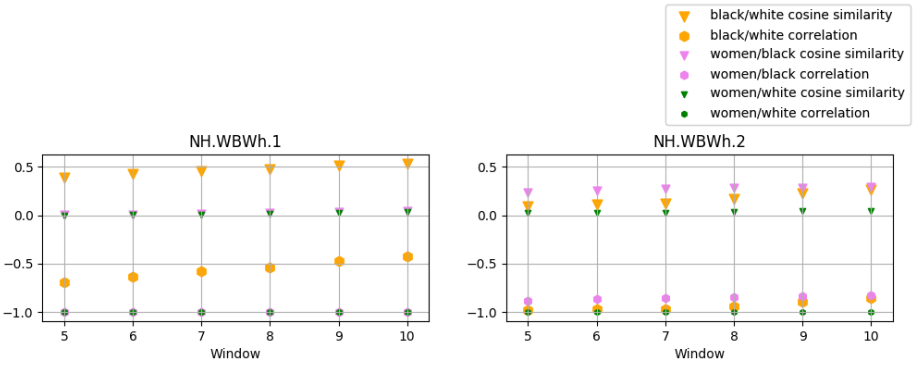

Fig. 2. Non-hate speeches NH.WBWh.1-2. In NH.WBWh.1, a 0.5 similarity score between *black* and *white* corresponds to a -0.5 value of anticorrelation, since the text is between *black* and *women*. NH.WBWh.1 - a pseudo-scientific argument on IQ, opposes *black* and *white*

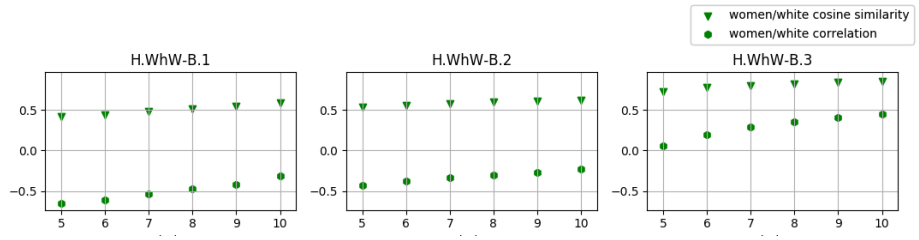

Fig. 3. Hate speeches H.WhW-B.1-3. In particular, in H.WhW-B.1-2 the keywords occur without a strong relation, whereas H.WhW-B.3 is explicitly on *white women*

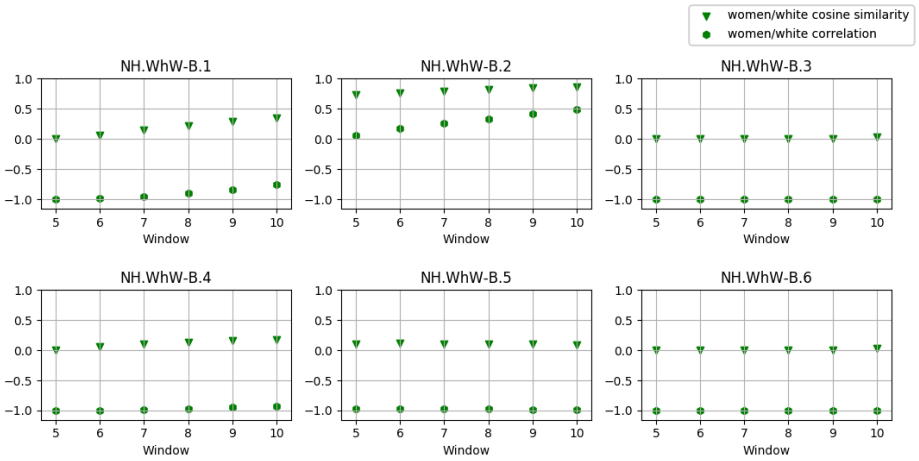

Fig. 4. Non-hate speeches NH.WhW-B.1-6. Five out of six documents show a maximal anticorrelation and a 0 similarity score. NH.WhW-B.2 is about *white women* ("Ivanka Voters").

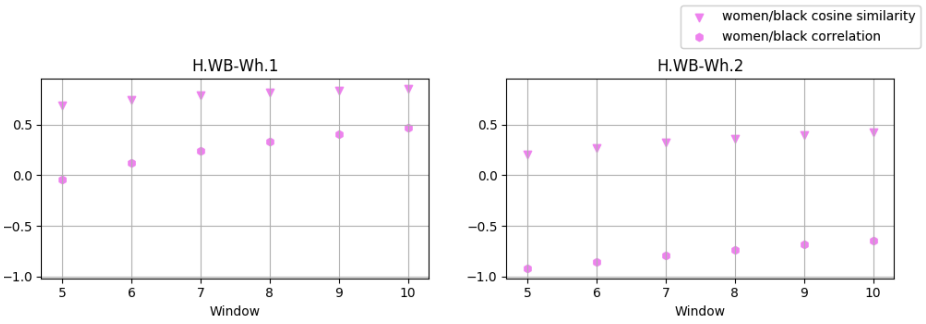

Fig. 5. Hate speeches H.WB-Wh.1-2. H.WB-Wh.1 is about *black women*, opposed to *black men*; H.WB-Wh.2 carries on an analogy between *blacks' rights* and *women's rights*

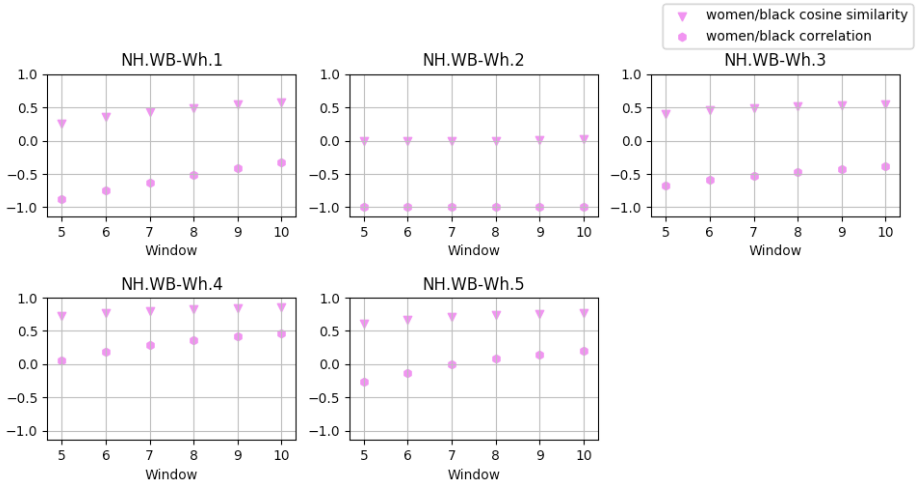

Fig. 6. Non-hate speeches NH.WB-Wh.1-5 show a great differentiation. In these cases, correlation is helpful to characterise more explicitly what does "similarity" mean

6 Conclusion

The paper shows how quantum correlation can clarify the less clear notion of similarity. The terms black and women can occur in hate-speeches with different relations, individuating *black women*, or opposing *women and blacks*. In particular:

- Low similarity values ($0 \leq cs \leq 0.5$) correspond to a maximum of anticorrelation ($-1 \leq r \leq$ -0.5) between the two stems. We have a double privative semantic opposition $-(A * B)$;
- Weak similarity ($0.5 \leq cs \leq 0.7$) correspond to weak anticorrelation or no correlation ($-0.5 \leq r \leq 0$);
- Higher similarity value ($0.7 < cs \leq 1$) corresponds to weak or strong correlations ($0 \leq r \leq 1$): $(A \leftrightarrow B)$;

The method seems promising as it concerns Digital Humanities and Machine Learning. Many machine learning techniques make use of human beings to label the corpus to avoid to define the involved labels, at the risk of mistakes and ambiguity. Quantum semantics offers a different insight on meaning, which can be useful to re-classify the corpus. For example, many hate speeches present strong anti-correlations between terms no matter of the width of the window. Furthermore, similar *semantic profiles* such as NH.WhW-B.2, H.WhW-B.3, NH.BW-Wh.4, H.BW-Wh.1 reveal a similar topic (black or white women) and show a sexist connotation, no matter how they have been labelled. In this paper we measured the correlation with reference to the single Pauli operator σ_x. In future works, measuring the expectations of Pauli's operators σ_y and σ_z, we could get an alternative way to measure entanglement [18], to be compared to Bell

inequalities used by Barros et al. [2]. On a similar line, the Born rule allows us to work on density matrices, giving an insight to the relation between meaning and Von Neumann information.

References

1. The online hate index (June 2018), http://www.adl.org/resources/reports/the-online-hate-index
2. Barros, J., Toffano, Z., Meguebli, Y., Doan, B.L.: Contextual query using bell tests. In: Atmanspacher, H., Haven, E., Kitto, K., Raine, D. (eds.) Quantum Interaction. pp. 110–121. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)
3. Bruza, P., Woods, J.: Quantum collapse in semantic space : interpreting natural language argumentation (06 2018)
4. Clauser, J.F., Horne, M.A., Shimony, A., Holt, R.A.: Proposed experiment to test local hidden-variable theories. Physical review letters **23**(15), 880 (1969)
5. Deleuze, G., Guattari, F.: A thousand plateaus: Capitalism and schizophrenia. Bloomsbury Publishing (1988)
6. Dubois, F., Toffano, Z.: Eigenlogic: A quantum view for multiple-valued and fuzzy systems. In: de Barros, J.A., Coecke, B., Pothos, E. (eds.) Quantum Interaction. pp. 239–251. Springer International Publishing, Cham (2017)
7. Eco, U.: From the tree to the labyrinth. Harvard University Press (2014)
8. Galofaro, F., Toffano, Z., Doan, B.L.: A quantum-based semiotic model for textual semantics. Kybernetes **47**(2), 307–320 (2018). https://doi.org/10.1108/K-05-2017-0187
9. Greimas, A.J., Rastier, F.: The interaction of semiotic constraints. Yale French Studies (41), 86–105 (1968), http://www.jstor.org/stable/2929667
10. Greimas, A.J.: Structural semantics: An attempt at a method. University of Nebraska Press (1983)
11. Lund, K., Burgess, C.: Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods, Instruments, & Computers **28**(2), 203–208 (Jun 1996). https://doi.org/10.3758/BF03204766
12. Melucci, M.: Introduction to information retrieval and quantum mechanics. Springer, Berlin, Heidelberg (2015)
13. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences). Cambridge university press (2004)
14. Petitot, J.: Morphogenesis of meaning. P. Lang (2004)
15. Rastier, F.: Sémantique interprétative. Presses universitaires de France (2009)
16. Singhal, A., et al.: Modern information retrieval: A brief overview. IEEE Data Eng. Bull. **24**(4), 35–43 (2001)
17. Song, D., Song, D., Bruza, P., Cole, R.: Concept learning and information inferencing on a high dimensional semantic space. In: ACM SIGIR 2004 Workshop on Mathematical/Formal Methods in Information Retrieval (MF/IR2004). Sheffield, United Kingdom, 25-29 July 2004. (2004). https://doi.org/10.1.1.370.4676
18. Susskind, L., Friedman, A.: Quantum mechanics: the theoretical minimum. Basic Books (AZ) (2014)
19. Van Rijsbergen, C.J.: The geometry of information retrieval. Cambridge University Press (2004)
20. Zinoviev, D.: Data Science Essentials in Python: Collect-Organize-Explore-Predict-Value. Pragmatic Bookshelf (2016)