

# Distributional techniques for philosophical enquiry

**Aurélie Herbelot**

Institut für Linguistik  
Universität Potsdam  
Karl-Liebknecht-Str. 24-25  
14476 Golm, Germany  
aurelie.herbelot@cantab.net

**Eva von Redecker**

Institut für Philosophie  
Humboldt Universität  
Unter den Linden 6  
10099 Berlin, Germany  
redecker@hu-berlin.de

**Johanna Müller**

Institut für Philosophie  
Humboldt Universität  
Unter den Linden 6  
10099 Berlin, Germany  
johannamueller@gmail.com

## Abstract

This paper illustrates the use of distributional techniques, as investigated in computational semantics, for supplying data from large-scale corpora to areas of the humanities which focus on the analysis of concepts. We suggest that the distributional notion of ‘characteristic context’ can be seen as evidence for some representative tendencies of general discourse. We present a case study where distributional data is used by philosophers working in the areas of gender studies and intersectionality as confirmation of certain trends described in previous work. Further, we highlight that different models of phrasal distributions can be compared to support the claim of intersectionality theory that ‘there is more to a phrase than the intersection of its parts’.

## 1 Introduction

Research in the social sciences rely heavily on linguistic analysis. Since what came to be called the ‘linguistic turn’ (Rorty, 1967) researchers across all humanities subjects have been highly aware of the fact that our access to the world, let alone cultural artefacts, is mediated by language and cast in our conceptual scheme.

Guided by the theory originating in Wittgenstein’s *Philosophical Investigations* (1953), one of the basic assumptions in contemporary analytic philosophy is now that the meaning of words is given by their usage in ordinary language. Conceptual analysis, i.e. the process of making explicit the rules that guide the applicability of a certain term, consequently forms a major occupation of the philosophical profession. The method exemplified by the French philosopher Michel Fou-

cault – discourse analysis – has become paradigmatic in the social sciences and cultural studies and constitutes a diachronic, historical version of the linguistic turn. One of the fundamental assumptions of this approach is that different eras produce different frameworks (or ‘episteme’ in Foucault’s terminology; see Foucault 1970) for understanding reality. Such frameworks manifest themselves as discursive patterns or specific formulations. According to Foucault, social power and the silencing of deviating utterances guarantee the temporary stability of a particular social regime. A similar methodology was pursued by the ‘Cambridge School’ of political theorists and historians, who tried to trace back the emergence of concepts like ‘state’ or ‘liberty’ not only to the ideas of a few canonical thinkers but to the ordinary use of those terms at the time (hence this approach is called ‘contextualism’, see Pocock 1975; Skinner 1998).

So far, such research has relied both on extensive manual work and on linguistic introspection. Manual methods, however, have clear drawbacks: they are time-consuming, expensive and likely to introduce bias in the data. This paper suggests that distributional techniques, as used in computational lexical semantics, may hold the key to automating the process of discourse analysis just described. We present a case study in philosophy, where two standard problems (the analysis of power in gender structures and the issue of so-called *intersectionality*) are reviewed in the light of distributional data. Not only do the produced distributions offer a rational way to highlight characteristic relationships between concepts, using an amount of data far greater than what could be annotated manually, but we show

that building on the relatively recent and novel research in composing distributions (Clark and Pulman, 2007), we can computationally illustrate the main thesis advocated by researchers on intersectionality.

This paper has a slightly unconventional format. We hope to exemplify a certain type of possible collaboration between computational linguistics and the humanities, which is less about providing an application to solve a particular problem than about drawing parallels between certain linguistic representations, known to have certain properties, and humanities-based theories. We felt that a fair amount of philosophical background was needed to show the relevance of our system's data to the particular type of investigation presented here. Therefore, a comprehensive philosophical introduction is given in §2. We then describe the system and corpus underpinning our research (§3 and 4) and discuss the theoretical aspects of lexical composition from a computational point of view, drawing parallels with the philosophical theory of intersectionality. Sections 5 and 6 discuss the worth of the data from a philosophical point of view.

## 2 Two philosophical problems

### 2.1 Gender and power

For a feminist philosopher, as for many people working in critical theory, the aim of research is twofold. First, to understand a given social structure, as for example the dynamics of gender relations and identities; second, to transform that structure towards greater freedom and equality. From its formation as an academic discipline in the second half of the 20th century onwards, one of the main concerns of feminist theory has been to show how social and institutional (man-made) factors have shaped what we are sometimes inclined to see as 'normal' or 'natural' gender identities. This approach is called social constructivism, because it ascribes the causal role for how gender identities emerge to processes of social construction. A prototypical example of this viewpoint, Simone de Beauvoir's famous claim that *one is not born a woman but becomes one* (Beauvoir, 1949) led to the distinction between biological 'sex' and social 'gender'. Such work has created an interest in the historical contingencies which decide what counts as a properly mascu-

line or feminine identity. But while competing biologicistic explanations of gender differences naturally have something they can point to (neurones, genes, or anatomy) social constructivist theories have sometimes lacked hard evidence for their claims. As a result, there has been a constant recurrence of theories claiming one natural cause for all aspects of gendered behaviour – and this, despite the fact that every single one has been proved wrong by the scientific community (Fine, 2010).

This state of affairs delineates a desideratum for feminist philosophy: giving more evidence of the cultural factors which instill gendered behaviour and associations in humans. While a lot of the cultural information concerning gender is visual (cinema, magazine covers, advertisement, etc), and the ways to initiating people into specific codes of gendered behavior can be non-verbal, the content of our notions of gender are vastly represented in text. Following this insight, Simone de Beauvoir reviewed a corpus of five modern novels to extract the characteristic aspects of what she coined the 'condition feminine' (Beauvoir, 1949). More recently, Judith Butler (1990) tried to explain how the use of certain concepts – gender, sex, desire, sexual practice – and associated notions were consolidating the dominant, binary distinction between masculinity and femininity.

The studies just mentioned, though central to their field, can always be met with suspicion. Objections such as *This is not how I use the word* or *You simply looked at books that distorted the understanding of the phenomenon in question* can only be met on the grounds of large-scale data. As a result, there have been attempts in historical research to produce statistical databases on gendered distributions. Hausen (1976), for example, focused on the turn of the 18th century in Germany, when a particular modern bourgeois understanding of gender roles is said to have emerged together with a new organisation of labour. However, the manual tasks involved in preparing such data are time-consuming and tedious and consequently, this type of work still has a limited coverage. We want to argue in this paper that entrusting the production of such resources to computational corpus linguistics would a) provide the ongoing philosophical investigation with an appropriate amount of data and b) help overcome the issues linked to the selective nature of the sources

a human reader might choose as relevant.

## 2.2 Intersectionality

Coined by the legal scientist Kimberlé Crenshaw (1991), the term **intersectionality** has spread into the humanities and social sciences as a perspective which has widened the scope of research on inequality and oppression within social contexts.

Scientists working with the intersectional perspective claim that the combination (intersection) of various forms of inequality (for example being black in a white dominated environment or being a woman in an environment dominated by men) makes a qualitative difference not only to the self-perception/identity of social actors, but also to the way they are addressed through politics, legislation and other institutions (Ngan-Ling Chow, 2011).

The founding case out of which Kimberlé Crenshaw developed the concept of intersectionality was a law suit that black women filed against the hiring policy of General Motors. Within the traditionally race and gender segregated automobile industry, women were only allowed to work in customer service or other office jobs while African-Americans were confined to factory work. As a consequence, African-American women faced the problem of being denied both office jobs and factory work. The women filed a lawsuit for discrimination. But the case was dismissed on the grounds that the plaintiffs hadn't been able to prove that they had been discriminated for either racial or sexist reasons. This case demonstrated our general additive understanding of discrimination: we act against sexism and we act against racism but we fail to address cases where they interact. The court was unable to take this interaction into account and Crenshaw made the case for a reform of the US anti-discrimination-law, which was based on the situation of white women in gender-dependent cases and of men in race-dependent cases.

So the aim of intersectionality is to localise and make visible discrimination that traditional thought cannot conceive of, in particular where various forms of oppression or inequality overlap. The claim is that social categories can only be explained in relation to other categories that define us as social beings within a society. For instance, what it means to be in the situation of a black woman can only be gauged in relation to

the political, cultural, socio-economic, religious, etc. background of that woman, and not with respect to being a woman or being a person of colour in isolation. Indeed, these different factors can override, exacerbate, conflict with each other, or simply run in parallel when they come to interact. Additionally, intersectionality also reminds us that the meaning of social categories also depends on historical eras. This proliferation of factors increases the complexity a researcher is confronted with to a point where it is doubtful whether the intersectional perspective can be transformed into a manageable methodology at all. Dealing with this level of complexity using automatically produced data might be of help on two levels:

1. the synchronic level, at which the intersectionality researcher seeks to grasp the qualitative differences between the concepts of, say, woman and black woman.
2. the diachronic level, at which historians working with intersectionality research conceptual change over texts from various eras.

In this work, we concentrate on the conceptual aspects of the synchronic level. In what follows, we will attempt to show that the intersectional approach can indeed be illustrated by the linguistic data obtained from a large contemporary corpus.

## 3 A distributional semantics system

### 3.1 Distributional semantics

Presented as a complement to model-theoretic semantics, distributional semantics aims to represent lexical meaning as a function of the contexts in which a given word appears (Wittgenstein, 1953; see also Harris, 1954, credited with the 'distributional hypothesis' which states that words which are similar in meaning occur in similar contexts).

Following this idea, some work in computational linguistics (starting with Harper, 1965) has been devoted to building and evaluating models which represent words as **distributions**, i.e., vectors in a multidimensional space where each dimension corresponds to a potential context for a lexical item (Curran, 2003). The notion of context itself has been studied to try and determine which representations work best for various tasks.

Word windows (Lund and Burgess, 1996), dependencies (Padó and Lapata, 2007) and syntactic relations (Grefenstette, 1994) have been proposed.

In our work, we use as context the words appearing in the same sentence as the query. This simple model is attractive from the point of view of using distributional techniques across the wide range of texts considered by humanities researchers. It ensures that a corpus is processable as long as it is digitalised – regardless of the language it is written in and the era it belongs to. Given that resources for parsing rare languages and older states of modern languages are still scarce, the word-based model has the advantage of flexibility.

Another issue in producing distributions relates to weighing the various dimensions. A number of possibilities have been suggested. Binary models attribute a weight of 1 to a context if it co-occurs at least once with the term that the distribution must represent, and 0 otherwise. Frequency-based models use as weights the number of co-occurrences of a particular context with the term under consideration. More complex models use functions like mutual information, which attempt to represent how ‘characteristic’ a particular context is for the term rather than how ‘frequent’ it is in conjunction with that term. The notion of a characteristic context is particularly important to us, as we wish to provide conceptual representations (distributions) which mirror what a ‘standard’ individual would associate with a given word. To achieve this, frequency models are not sufficient. Words like *do*, *also*, *new*, etc co-occur with many terms but are in no meaningful relation with those terms. Instead, we want to choose a function which gives high weights to contexts that appear frequently with the term to be modelled and not very frequently with other terms. By doing this, we will have a way to describe salient associations for a particular concept. In §3.3, we will spell out such a function, borrowed from Mitchell and Lapata (2010).

### 3.2 Intersectionality in linguistic terms

It has been suggested that in order to integrate distributional semantics with model theoretic formalisms, methods should be found to compose the distributions of single words (Clark and Pulman, 2007). It is clear that the representation of *carnivorous mammal* in formal semantics can be

written as  $\text{carnivorous}'(x) \wedge \text{mammal}'(x)$  but it is less clear how the lexical semantics of the phrase should be described in distributional terms.

The work done so far on distributional compositionality has focused on finding equivalents for the well-known formal semantics notion of intersection. All models assume that the intersective composition of two elements should return a distribution, i.e. a lexical meaning, which is made of the individual distributions, or meanings, of those elements. But there are differences in how those models are evaluated. Two categories can be drawn: models designed to emulate the distribution of the resulting phrase itself, as it would be observed given a large enough corpus (Guevara, 2010 and 2011; Baroni and Zamparelli, 2010), and those which only focus on the composition operation and try to produce an adequate representation of the semantic intersection of the phrase’s components, independently from the phrasal distribution (Mitchell and Lapata, 2010; Grefenstette and Sadrzadeh, 2011). The former, which we will refer to as **phrasal models** are trained and evaluated against phrases’ distributions while the latter, **intersective models**, call for task-based evaluations (for instance, similarity ratings: see Mitchell and Lapata, 2010).

We argue that phrasal and intersective models are bound to produce different aspects of meaning. Consider, for instance, the phrase *big city*. Principles of semantic intersection tell us that a big city is a city which is big. This is a correct statement and one which should come out of the composition of the *big* distribution and the *city* distribution<sup>1</sup>. But arguably, there is more to the meaning of the phrase (see Partee, 1994, for a discussion of non-intersective adjectival phrases). We expect people to readily associate concepts like *loud*, *underground*, *light*, *show*, *crowd* to the idea of a big city. Our hypothesis is that this ‘extra’ (non-intersective) meaning can be clearly observed in **phrasal distributions** while it is, in some sense, ‘hidden’ in distributions which are the result of a purely intersective operation (because the entire distributions of the two components are used, and not just the contexts relevant to the particular association of *big* and *city*).

This observation, made at the linguistic level,

<sup>1</sup>For the sake of this argument, we will ignore the suggestion that the gradable adjective might affect the intersective status of the phrase

is also the foundation of intersectionality theory. The argument, presented in Section 2, is that the prejudices attached to black women, for instance, (that is, the way that the concept *black woman* is understood and used) are different from the simple combination of the prejudices attached to black people and women separately. So although it is correct to say that a black woman is a woman who is black (in the relevant sense of *black*), the concept reaches further.

If the basic tenet of intersectionality theory holds, and if we accept that distributions are a valuable approximation of lexical meaning, we would expect that the **phrasal distribution** of, say, *black woman* would significantly differ from its **compositional distribution**. Further, such a significant difference would also have linguistic relevance, as it would indicate the need to take phrasal distributions into account when ‘computing meaning’ via distributional techniques. Both phrasal models and intersective models could be said to contribute to a complete and accurate representation.

In Section 6, we will make a first step in investigating this issue by thoroughly analysing the phrasal and compositional distributions of *black woman*.

### 3.3 System description

The two systems we use in this paper have the same basis. They produce a distribution for a phrase based on a raw Wikipedia<sup>2</sup> snapshot, pre-processed to remove the wiki markup (Flickinger et al, 2010). Distributions are vectors in a space  $S$  made of 10000 dimensions which correspond to the 10000 most frequent words in the corpus. The distribution of a word or phrase in the corpus is taken to be the collection of all words that co-occur with that word or phrase within a single sentence (we use a list of stop words to discard function words, etc). The weight of each co-occurring term in the distribution is given by a function borrowed from Mitchell and Lapata (2010):

$$w_i(t) = \frac{p(c_i|t)}{p(c_i)} = \frac{freq_{c_i,t} * freq_{all}}{freq_t * freq_{c_i}} \quad (1)$$

where  $w_i(t)$  is the weight of word  $t$  for dimension  $i$ ,  $freq_{c_i,t}$  is the count of the co-occurrences of a

<sup>2</sup><http://www.wikipedia.org/>

context word  $c_i$  with  $t$ ,  $freq_{all}$  is the total number of words in the corpus,  $freq_t$  and  $freq_{c_i}$  are respectively  $t$  and  $c_i$ ’s frequencies in the corpus.

We choose an intersective model based on multiplication, as this operation has been shown to give excellent results in previous experiments (Erk and Padó, 2008; Mitchell and Lapata, 2010): the distributions of the two components of the phrase are multiplied in a point-wise fashion to give the final distribution. This corresponds to the model  $\mathbf{p}=\mathbf{u}\odot\mathbf{v}$  of Mitchell and Lapata.

As for the phrasal model, the final distribution is simply the distribution obtained from looking at the co-occurrences for the phrase itself.

The data passed on to the philosophers for further consideration takes the form of a list of the 100 most ‘characteristic’ contexts for the query, that is, the 100 words with highest weights in the distribution, filtered as described in §3.3.1.

#### 3.3.1 Filtering the results

One potential issue with our implementation is that words belonging to frequent named entities end up in the top characteristic contexts for the query. So for instance, *wonder* is one of the most characteristic contexts for *woman* because of the comic character *Wonder Woman*. Arguably, such contexts should only be retrieved if they are as significant in their ‘non-name’ form as in their named entity form (e.g. if the string *wonder woman* was significantly more frequent than *Wonder Woman*, there would be a case for retaining it in the results). We filter the relevant named entities out using the following heuristic:

1. We call the query  $q$  and its capitalised version  $Q$ . Let  $c_1\dots c_n$  be the top characteristic contexts for  $q$  and  $C_1\dots C_n$  their capitalised equivalent.
2. For each context  $c_k$  in  $c_1\dots c_n$ :
  - (a) We compute the corpus frequencies of the patterns  $qc_k$ ,  $c_kq$ ,  $qwc_k$  and  $c_kwq$ , where  $w$  is an intervening word.  $s_k$  is the sum of those frequencies.
  - (b) Similarly, we compute the corpus frequencies of the patterns  $QC_k$ ,  $C_kQ$ ,  $QwC_k$  and  $C_kwQ$ , where  $w$  is an intervening word.  $S_k$  is the sum of those frequencies.

- (c)  $r$  is the ratio  $S_k/s_k$ . If  $r$  is over a certain threshold  $t$ , we remove  $c_k$  from the characteristic contexts list.

In our experiments, we use  $t = 0.6$ . This threshold allows us to successfully remove many spurious contexts. For example, *wonder* and *spider* are deleted from the results for *woman* (among others) and *isle*<sup>3</sup>, *elephant* and *iron* from the results for *man*.

#### 4 The corpus

The number of experiments that can be devised using distributional techniques is only limited by the number of digitalised corpora available to researchers in the humanities. It is easy to imagine a range of comparative studies showing the conceptual differences highlighted by the use of a word or phrase at various times, in various countries, or in various communities. The aim of this study is to analyse the discursive use of some concepts over a fairly large sample. We chose the English Wikipedia<sup>4</sup> as our corpus because of its size and because of its large contributor base (around 34,000 ‘active editors’ in December 2011<sup>5</sup>). Wikipedia’s encyclopedic content also makes it less explicitly biased than raw Internet text, although we have to be aware of implicit bias: most of Wikipedia’s contributors are male and the encyclopedia’s content is heavily skewed towards items of popular culture (Cohen, 2011). The latter point is unproblematic as long as it is acknowledged in the discussion of the results.

In the next two sections, we analyse the distributions obtained for the phrases *man*, *woman*, *black woman* and *Asian woman*. It is worth mentioning that more data was extracted from our corpus, which, due to space constraints, we will not discuss here. The broad claims made with respect to the above four noun phrases, however, are consistent with the rest of our observations.

#### 5 Discussing gender

This section discusses the produced distributions from the perspective of gender theory. The aim of the discussion is to illustrate the type of

information that may be relevant for discourse analysis. Note that it follows the philosophical methodology highlighted in Haslanger (2005) for conceptual analysis.

Table 1 shows the most characteristic contexts for *woman* and *man*, after filtering. There are three levels on which the data discussed here could be usefully interpreted within feminist research – the conceptual, the constructivist and the deconstructivist. We will concentrate on the first two.

In recent years a strong trend in Gender Studies has emphasised that our investigations shouldn’t repeat the historical bias which regards men as ‘universal’ (or default) and women as ‘particular’, as a specific ‘other’ to be investigated (Honegger, 1991). An advantage of our automatically produced data is that it returns just as much material to focus on the cultural fabrication of masculinity as on that of femininity. The male position proves to indeed carry a broader variety of seemingly non-gendered contexts, for instance, *wise*, *innocent*, *sin*, *fat*, *courage*, *salvation*, *genius*, *worthy* and *rescued* – none of which is characteristic of *woman*. But what is most striking is the strong occurrence of military contexts. We find *enlisted* at the top of the list, followed by *wounded*, *IRA*, *officers*, *militia*, *regiments*, *garrison*, *platoons*, *casualties*, *recruits* and diverse military ranks. It is sometimes unclear what counts as ‘military-related’ (see *killing*, *brave*). We would have to go back to the original text to investigate this. But we see here very clearly how attributes that rank high when it comes to defining stereotypical masculinity and might be thought as ‘general’ characteristics clearly owe their prominence to the military cluster. The characteristic contexts list seems to give distilled evidence to what has as yet still only been partly analyzed in socio-historical research, namely how the norms of masculinity are to a large extent of military descent (for the German context see Frevert, 2001). *Brave*, *angry*, *courage*, *cruel* are all things that Wikipedia – just like popular imagination – won’t associate with women.

The meaning of *woman* seems to revolve around the three interrelated clusters of reproduction, sex and love. *Pregnant* and *pregnancy* rank very high, as well as reproduction-related terms such as *abortion*, *children* and *mothers*. There are more sexual terms (*sexually*, *sexual*, *sexual-*

<sup>3</sup>As in *Isle of Man*

<sup>4</sup>[http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)

<sup>5</sup><http://stats.wikimedia.org/EN/SummaryEN.htm>

Woman	Man
women, woman, pregnant, feminist, abortion, womens, men, husbands, elderly, pregnancy, sexually, rape, breast, gender, equality, minorities, lesbian, wives, beautiful, attractive, pornography, dressed, sexual, marry, sexuality, dress, est., wear, young, sex, african-american, naked, comfort, homosexual, discrimination, priesthood, womens, violence, loved, children, clothes, man, male, marriages, hair, mysterious, wearing, homeless, loves, boyfriend, wore, her., ladies, mistress, lover, attitudes, hiv, advancement, relationships, homosexuality, wealthy, mothers, worn, murdered, ordained, mortal, unnamed, girls, depicts, slavery, lonely, female, equal, cancer, goddess, roles, abuse, kidnapped, priests, portrayal, witch, divorce, screening, clothing, murders, husband, romantic, forbidden, loose, excluded	men, man, enlisted, women, wise, homosexual, wounded, gay, woman, dressed, young, elderly, ira, homeless, wives, brave, angry, officers, marry, marched, sexually, wealthy, killed, wounds, innocent, militia, homosexuality, mans, mysterious, god, tin, elves, mortal, ladies, wearing, priesthood, sin, con, courage, fat, equality, numbering, regiments, garrison, numbered, brotherhood, murdered, rape, lonely, platoon, casualties, knew, recruits, reinforcements, recruited, blind, loved, sexual, sex, thousand, mask, clothes, salvation, commanded, loves, lover, sick, detachment, genius, cruel, gender, killing, col., lt., drunk, worthy, tall, flank, convicted, surrendered, contingent, rescued, naked

Table 1: Most characteristic contexts for *woman* and *man*, after filtering

ity, sex) in the characteristic list for *woman* and mentions of *loved*, *loves*, *lover* are higher up than in the results obtained for *man*. Further, a variety of terms, mostly absent from the *man* list, create a close link between femininity and relationality: *husband(s)*, *marriage* (though, further down, *divorce* comes up too), *boyfriend* and *relationships*. While *beautiful*, *attractive*, *comfort* and *romantic* might at least suggest that positive sentiments are attached to the inbuilt feminine relationality, another set of female contexts highlights the very vulnerability inscribed in the cluster around intimacy: *rape*, *pornography*, *violence*, *slavery*, *abuse*, *kidnapped* quantitatively capture a correlation between relationality, sexuality and violence which characterises the use of the lexeme *woman*.

Another set of exclusively feminine concepts which at first sight seem to create interesting singular contexts – *breast*, *comfort*, *hair*, *HIV*, *cancer* – are united by reference to a physicality that seems, apart from the wounds apparently contracted in war, peculiarly absent in man. Such clustering sheds light on the fact that certain associations ‘stick’ to women and not to men. Though it takes two to marry or divorce and have children, those exclusively form contexts for woman. Most dramatically, this can be observed when it comes to rape. Though the majority of cases im-

ply a male perpetrator, *rape* is very high up, in 12th position, in the female list (that is before any mention of love), while it is returned as characteristic of men only to the extent that *loneliness* or *brotherhood* are, at rank 49.

These observations highlight the kind of associations implicitly present in discursive data – whether retrieved by machines or humans. We do not learn how matters are ‘in fact’ but simply integrate the linguistic patterns most characteristic for a certain phenomenon. This, again, does have tremendous effects on reality – so-called ‘constructive’ effects. Indeed, when it comes to phenomena that touch on human self-understanding, discourse implies more than a mirror of meaning. It partakes in the making of real identities. It provides the material people craft their self-understanding from and model their behavior after. It is this very effect of our characteristic usage of language which prompts social philosophers to ascribe ‘power’ to language.

## 6 Discussing intersectionality

Table 2 shows the most characteristic contexts for the phrase *black woman* after filtering, as given by the interjective and phrasal models. We should point out that the phrase *black*

Multiplicative model	Phrasal model
stripes, makeup, pepper, hole, racial, white, woman, spots, races, women, whites, holes, colours, belt, shirt, african-american, pale, yellow, wears, powder, coloured, wear, wore, colour, dressed, racism, leather, colors, hair, colored, trim, shorts, silk, throat, patch, jacket, dress, metal, scarlet, worn, grey, wearing, shoes, purple, native, gray, breast, slaves, color, vein, tail, hat, painted, uniforms, collar, dark, coat, fur, olive, bear, boots, paint, red, lined, canadiens, predominantly, slavery	racism, feminist, women's, slavery, negro, ideology, tyler, filmmaker, african-american, ain't, elderly, whites, nursing, patricia, abbott, gloria, freeman, terrestrial, shirley, profession, julia, abortion, diane, possibilities, argues, reunion, hiv, blacks, inability, indies, sexually, giuseppe, perry, vince, portraits, prevention, beacon, gender, attractive, tucker, fountain, riley, beck, comfortable, stern, paradise, twist, anthology, brave, protective, lesbian, domestic, feared, breast, collective, barbara, liberation, racial, rosa, riot, aunt, equality, rape, lawyers, playwright, white, argued, documentary, carol, isn't, experiences, witch, men, spoke, slaves, depicted, teenage, photos, resident, lifestyle, aids, commons, slave, freedom, exploitation, clerk, tired, romantic, harlem, celebrate, quran, interred, star-gate, alvin, ada, katherine, immense

Table 2: Most characteristic contexts for *black woman*. Multiplicative and phrasal model, after filtering

*woman/women* only occurs 384 times in our corpus, so the vector obtained through the phrasal model suffers from some data sparsity problems.<sup>6</sup> In particular, overall infrequent events are given high weights by our algorithm, resulting in a relatively high number of surnames being present in the produced vector. Despite this issue, a number of observations can be made, which agree with both our linguistic and philosophical expectations.

We first considered to what extent the phrasal and multiplicative models emphasised the characteristics already present in their components' distributions. We found that the top contexts for the phrasal distribution of *black woman* only overlap 17 times with the top contexts of *woman* and 9 times with the top contexts for *black*. The multiplicative model produces an overlap of only 12 items with *woman* but 64 with *black*.<sup>7</sup> This highlights a large conceptual differ-

<sup>6</sup>We should add that this small number of occurrences is in itself significant, and mirrors problems of social marginalisation. We note that the phrase *African-American woman/women* is even sparser, with 236 occurrences in our corpus.

<sup>7</sup>The weights in the *black* vector clearly override those from the *woman* vector. Mitchell and Lapata (2010) discuss possible improvements to the multiplicative model which involve putting a positive weight on the noun component when performing the composition.

ence between the phrase seen as a single entity and its components. In contrast, the composition of the constituents via the multiplicative model returns a distribution fairly close to the distribution of those constituents.

We looked next at the characteristic contexts that were particular to each representation. We found that the phrasal model vector presents 73 terms which are absent from the top contexts for *black* and *woman*. In contrast, none of the terms in the top contexts of the multiplied vector is specific to the composed phrase: all of them are either highly characteristic of *black* or *woman* (e.g. *racial*, *African-American* and *breast*, *dress*).<sup>8</sup> This indicates that the salient contexts for the phrase are very different from the associations commonly made with its constituents.

Finally, we observed that the vector obtained through the phrasal model only overlaps 8 times with the composed one. The shared words are *racial*, *white*, *whites*, *African-American*, *racism*, *breast*, *slaves* and *slavery*. Again, we can conclude from this that the two representations capture very different aspect of meaning, although they both retrieve some high-level associations with the concept of race and race-related history.

<sup>8</sup>Because our system does not perform any sense disambiguation, we return contexts such as *pepper* for *black pepper*, *hole* for *black hole*, etc.



From the point of view of intersectionality, our results confirm the basic claim of the theory: there are cases where the discourse on individuals belonging to two different social groups is radically different than the discourse pertaining to those social groups taken separately.

In addition, the data supports further arguments made by intersectionality researchers. In particular, comparing the distributions for *woman*, *black woman* and *Asian woman*<sup>9</sup> shows that colour or ethnicity has a crucial impact on how women are represented. Looking at *woman*, the word *rape* appears at position 12, but it appears much further down the list in *black woman* and not at all in *Asian woman*. At the same time the word *nursing* is only associated with black women while *pornography* hits position number 3, shortly followed by *exotic* and *passive* when we look at *Asian woman*. These three words do not occur in the top contexts for *woman* or *black woman*. This indicates that we are getting results concerning sexuality which depend on ‘ethnicity’ connotation: white women are shown as victims of abuse (*rape*), black women as responsible for *nursing*, and Asian women represented as *passive* and objects of *pornography*.

Just looking at this data (and there would be a lot more to analyse) we can find a connection with what the latest historical research working with intersectionality has brought to light: historians have shown that the historical discourse on prostitution has increased and reinforced racist stereotypes and prejudices. Whyte (2012) shows, (looking at the ‘white slavery’ panic of the early twentieth century which is a key point in the history of prostitution) that the construction of the white, innocent victim of prostitution is central to the creation of the myth of the ‘white slavery’ in many ways and that it has shaped the construction and understanding of contemporary human trafficking. The broader history of slavery (particularly in the North American context) forms the backdrop for ‘writing out’ women of colour as victims of sexual violence. This is appropriately illustrated by our data.

---

<sup>9</sup>Due to space constraints, we are not showing the distribution of *Asian woman*.

## 7 Conclusion

This paper sought to demonstrate that linguistic representations of the type used in distributional semantics may provide useful data to humanities researchers who analyse discursive trends. We presented a case study involving two subfields of philosophy: gender theory and intersectionality.

We hope to have shown that a) distributional data is a useful representation of social phenomena which have been described by theorists and social scientists but never linguistically observed on a large scale b) this data lends itself to a fine-grained analysis of such phenomena, as exemplified by the discussions in §5 and §6. Further, we have highlighted that the philosophical theory of intersectionality can be illustrated, at least for some concepts, via a quantitative analysis of the output of different distributional models. We suggest that this observation should be investigated further from the point of view of computational linguistics: there may be some aspect of meaning which is not expressed by those distributional compositional models that do not take phrasal distributions into account (i.e. additive, multiplicative, circular convolution models).

A natural extension of this work would be to design experiments focusing on particular types of discourse and corpora, and pursue conceptual analysis at the diachronic level. This presupposes the existence of digitalised corpora which may not be available at this point in time. Efforts should therefore be made to acquire the needed data. We leave this as future work.

## Acknowledgments

This work was in part enabled by an Alexander von Humboldt Fellowship for Postdoctoral Researchers awarded to the first author. We would like to thank the Alexander von Humboldt Foundation for their financial support.

## References

- Marco Baroni and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP10).
- Simone de Beauvoir. 1949. *Le deuxième sexe*. Gallimard, Paris.

- Judith Butler. 1990. *Gender Trouble. Feminism and the Subversion of Identity*. Routledge, New York.
- Stephen Clark and Stephen Pulman. 2007. Combining Symbolic and Distributional Models of Meaning. *Proceedings of the AAAI Spring Symposium on Quantum Interaction*, pp.52–55. Stanford, CA.
- Noam Cohen. 2011. Define Gender Gap? Look Up Wikipedias Contributor List. *The New York Times*, 31 January 2011, pp.A1 New York edition.
- Kimberlé Crenshaw. 1991. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, 43:6, pp. 1241–1299.
- James Curran. 2003. *From Distributional to Semantic Similarity*. PhD dissertation. Institute for Communicating and Collaborative Systems. School of Informatics. University of Edinburgh, Scotland, UK.
- Katrin Erk and Sebastian Padó. 2008. A Structured Vector Space Model for Word Meaning in Context. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, HI.
- Cordelia Fine. 2008. *Delusions of Gender: The Real Science Behind Sex Differences*. Icon Books. London.
- Dan Flickinger and Stephan Oepen and Gisle Ytrestøl. 2010. WikiWoods: Syntacto-Semantic Annotation for English Wikipedia. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC10)*.
- Michel Foucault. 1966. *Les mots et les choses*. Gallimard, Paris.
- Ute Frevert. 2001. *Die kasernierte Nation. Militärdienst und Zivilgesellschaft in Deutschland*. Beck, Munich.
- Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011. Experimental Support for a Categorical Compositional Distributional Model of Meaning. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 1394–1404. Edinburgh, Scotland, UK.
- Gregory Grefenstette. 1994. *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers.
- Emiliano Guevara. 2010. A Regression Model of Adjective-Noun Compositionality in Distributional Semantics. *Proceedings of the 2010 Workshop on Geometrical Models of Natural Language Semantics, ACL 2010*, pp. 33–37. Uppsala, Sweden.
- Emiliano Guevara. 2011. Computing Semantic Compositionality in Distributional Semantics. *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*, pp. 135–144. Oxford, England, UK.
- Kenneth E. Harper. 1965. Measurement of similarity between nouns. *Proceedings of the 1965 conference on Computational linguistics (COLING 65)*, pp. 1–23. Bonn, Germany.
- Zelig Harris. 1954. Distributional Structure. *Word*, 10(2-3):146–162.
- Sally Haslanger. 2005. What Are We Talking About? The Semantics and Politics of Social Kinds *Hypatia*, 20:10–26.
- Karin Hausen. 1976. Die Polarisierung der “Geschlechtscharaktere”. Eine Spiegelung der Dissoziation von Erwerbs- und Familienleben. *Sozialgeschichte der Familie in der Neuzeit Europas*. Conze and Werner, Editors, pp. 363–393 Klett-Cotta, Stuttgart.
- Claudia Honegger. 1991. *Die Ordnung der Geschlechter: Die Wissenschaften vom Menschen und das Weib, 1750/1850*. Campus, Frankfurt am Main and New York.
- Kevin Lund and Curt Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers*, 28, pp. 203–208.
- Jeff Mitchell and Mirella Lapata. 2010. Composition in Distributional Models of Semantics. *Cognitive Science*, 34(2010):1388–1429.
- Esther Ngan-Ling Chow. 2011. *Analyzing Gender, Intersectionality, and multiple inequalities: global, transnational and local contexts*. Esther Ngan-Ling Chow, Marcia Texler Segal and Tan Lin, Editors. Emerald Group Publishing.
- Sebastian Padó and Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2), 161–199.
- Barbara Partee. 1994. Lexical Semantics and Compositionality *Invitation to Cognitive Science, second edition. Part I: Language*. Daniel Osherson, General Editor. Lila Gleitman and Mark Liberman, Editors. MIT Press.
- John G.A. Pocock. 1975. *The Machiavellian moment. Florentine political thought and the Atlantic republican tradition*. Princeton University Press, Princeton, NJ.
- Richard Rorty. 1967. *The Linguistic Turn: Recent Essays in Philosophical Method*. Richard Rorty, Editor. University of Chicago Press, Chicago.
- Quentin Skinner. 1998. *Liberty before liberalism*. Cambridge University Press, Cambridge, UK.
- Dominic Widdows. 2008. Semantic Vector Products: Some Initial Investigations. *Second AAAI Symposium on Quantum Interaction*. Oxford, UK.
- Christine Whyte. 2012. Praise be, prostitutes as the women we are not – Using intersectionality to analyse race, class and gender in history. *Intersektionalität und Kritik*. Vera Kallenberg, Jennifer Meyer, Johanna M. Müller, Editors. Wiesbaden. Forthcoming.
- Ludwig Wittgenstein. 1953. *Philosophische Untersuchungen*. Blackwell, Oxford, UK.