Distributional techniques for philosophical enquiry

Aurelie Herbelot, Eva von Redecker and Johanna M. Müller

1Universität Potsdam
Institut für Linguistik

2Humboldt Universität zu Berlin
Institut für Philosophie

LATECH 2012
Outline

1. Introduction
2. Genders and concepts
3. Distributional semantics
4. Intersectionality in linguistic terms
5. System description
6. Investigating concepts using distributions
   - Gender studies
   - Intersectionality
7. Conclusion
Discourse analysis in philosophy

The ‘linguistic turn’ (Rorty, 1967)

Our access to the world is mediated by language and cast in our conceptual scheme.

- Discourse analysis: investigate conceptual framework of a certain place at a certain time by looking at how people speak/spoke (see Foucault, 1970).
- Lots of time-consuming work in archives (studious philosopher) or lots of self-centered introspection (lazy philosopher).
Introduction

Issues in discourse analysis

- Manual methods are time-consuming, expensive and likely to introduce bias in the data.

A new tool?

Distributional techniques, as used in computational lexical semantics, may help automate the process of discourse analysis.

- 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.
What are gender studies?

Arnie
(Biological man)

Claudia Schiffer
(Biological woman)
What are gender studies?

Mary
(Biological man)

Patti Smith
(Biological woman)
What are gender studies?

- Is there a male/female binary? If so, where does it come from?
- What to do (conceptually) with Mary and Patti?
- Medical research has shown that (biological) gender was linear.
- **Social constructivism** ascribes the causal role for how binary gender identities emerge to processes of social construction.
Our aim

Help the studious philosopher

We want to produce conceptual representations which reflect the ‘discourse’ of a certain time and place and give the output to philosophers for further analysis.
Lexical meaning is a function of the contexts in which a given word appears (Harris, 1954).

Words represented as **distributions**, i.e., vectors in a multidimensional space where each dimension corresponds to a potential **context** for a lexical item (Curran, 2003).

E.g. concept *cat* represented as a vector with high values along the dimensions *mouse*, *meow*, *sleep*, etc.:

\[
\text{cat} \equiv \text{new (0.01) city (0.005) ... mouse (0.8) ... meow (0.9) business (0.0003) ... sleep (0.7) dog(0.6)...}
\]

But what about phrases? (*sleepy cat*)
Kimberlé Crenshaw (1991)

The combination (intersection) of various forms of inequality 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.

- Founding case: a law suit that African American women filed against the hiring policy of General Motors (DeGraffenreid v. General Motors, 1977).
- Crenshaw made the case for a reform of the US anti-discrimination-law.
- Her work was further influential in the drafting of the equality clause in the South African Constitution.
- The concept black woman is not the addition of black and woman.
Outline

1. Introduction
2. Genders and concepts
3. Distributional semantics
4. Intersectionality in linguistic terms
5. System description
6. Investigating concepts using distributions
   - Gender studies
   - Intersectionality
7. Conclusion
Composing distributions

- Integrate distributional semantics with model theoretic formalisms: compose the distributions of single words (Clark and Pulman, 2007). The intersective composition of two elements should return a distribution.

- Two categories:
  - 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). Trained and evaluated against phrases’ distributions.
  - models which only focus on the composition operation, independently from the phrasal distribution (Mitchell and Lapata, 2010; Grefenstette and Sadrzadeh, 2011). Task-based evaluation.

- We call the former **phrasal models** and the latter, **intersective models**.
The meaning of phrases

Is intersection enough?

A *big city*: just a city which is big?
See *loud, underground, advertisement, crowd, Phantom of the Opera*...

- There is more to composition than intersection (see Partee, 1994).
- Our hypothesis: there is ‘extra’ (non-intersective) meaning which can be clearly observed in **phrasal distributions** and which is ‘hidden’ in distributions that are the result of a purely intersective operation.
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 **intersective distribution**.
Outline

1. Introduction
2. Genders and concepts
3. Distributional semantics
4. Intersectionality in linguistic terms
5. System description
6. Investigating concepts using distributions
   - Gender studies
   - Intersectionality
7. Conclusion
System description

Basic system implementation

- Raw Wikipedia snapshot, preprocessed 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: 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):

$$w_i(t) = \frac{p(c_i|t)}{p(c_i)} = \frac{freq_{c_i,t} \cdot freq_{all}}{freq_t \cdot freq_{c_i}}$$  \hspace{1cm} (1)
The intersective and phrasal models

- The intersective model is based on multiplication: 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 \( p = u \odot v \) of Mitchell and Lapata.

- 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 to remove spurious named entities (Elephant Man, Spider Woman)...
Outline

1. Introduction
2. Genders and concepts
3. Distributional semantics
4. Intersectionality in linguistic terms
5. System description
6. Investigating concepts using distributions
   - Gender studies
   - Intersectionality
7. Conclusion
### Woman and Man

<table>
<thead>
<tr>
<th>Woman</th>
<th>Man</th>
</tr>
</thead>
<tbody>
<tr>
<td>women, woman, pregnant, feminist, abortion, women’s, 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</td>
<td>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</td>
</tr>
</tbody>
</table>

**Table:** Most characteristic contexts for *woman* and *man*, after filtering
Analysis of the *man* and *woman* distributions

- In *man*, predominance of military-related contexts: *enlisted*, *wounded*, *IRA*, *officers*, *militia*, *regiments*, *garrison*, *platoons*, *casualties*, *recruits*, etc. (Frever, 2001)

- The meaning of *woman* seems to revolve around the three interrelated clusters of reproduction, sex and love.
  - *pregnant*, *pregnancy*, *abortion*, *children*, *mothers*
  - *sexually*, *sexual*, *sexuality*, *sex*
  - *loved*, *loves*, *lover* (much higher up than in the results obtained for *man*)
Comparing distributions

- Certain associations ‘stick’ to women and not to men, and vice-versa.
  - Although it takes two to marry or divorce and have children, those are exclusively characteristic of women.
  - Although, women can in principle be *brave*, *angry*, *courage*, *cruel*, those are exclusively characteristic of men.
  - Although the majority of cases imply 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.
The distribution of *black woman*

<table>
<thead>
<tr>
<th>Multiplicative model</th>
<th>Phrasal model</th>
</tr>
</thead>
<tbody>
<tr>
<td>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</td>
<td>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, diaone, 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, stargate, alvin, ada, katherine, immense</td>
</tr>
</tbody>
</table>

**Table**: Most characteristic contexts for *black woman*. Multiplicative and phrasal model, after filtering
Quality of the distributions

- The phrase *black woman/women* only occurs 384 times in our corpus. Count for *African-American woman/women* = 236 occurrences.
- The vector obtained through the phrasal model suffers from some data sparsity problems.
- Overall infrequent events are given high weights by our algorithm, resulting in particular in a relatively high number of surnames being present in the produced vector.
- The data sparsity problem mirrors problems of social marginalisation.
Quantitative analysis

- Top contexts for the phrasal distribution of *black woman* overlap 17 times with the top contexts of *woman* and 9 times with the top contexts for *black*.
- Multiplicative model produces an overlap of only 12 items with *woman* but 64 with *black*. (The weights in the *black* vector clearly override those from the *woman* vector.)
- The phrasal model vector presents 73 terms which are absent from the top contexts for *black* and *woman*. In contrast, the multiplied vector returns no new information.
Confirmation of intersectionality theory

- Short summary: we observe a large conceptual difference 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.

- Our results confirm the basic claim of intersectionality 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.
Some more claims illustrated...

- Intersectionality theory claims that colour or ethnicity has a crucial impact on how women are represented.
- We compare the distributions for woman, black woman and Asian woman.
- 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.
- 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.
Outline

1. Introduction
2. Genders and concepts
3. Distributional semantics
4. Intersectionality in linguistic terms
5. System description
6. Investigating concepts using distributions
   - Gender studies
   - Intersectionality
7. Conclusion
Conclusion

- 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.
- Intersectionality as a theory is relevant for linguistics: it reminds us that the meaning representation of phrases should include more than the meaning of their components.
- In return, intersectionality can be illustrated via a quantitative analysis of the output of different distributional models.