Big Data, Big Meaning
Using distributional semantics in linguistic research

Florent Perek
University of Birmingham
Natural language processing (NLP)

Programming computers to use human language

"Ok Google"
Natural language processing (NLP)

- NLP is everywhere
- Fast change that happened over the last 10-15 years
  - Increasingly advanced statistical processing
  - Big Data

\[
12765 \times 964 = ?
\]
NLP and linguistics

- NLP has produced many techniques to process large amount of data and extract linguistic information from it
- Linguistic research can benefit a lot from these techniques
- Case in point: distributional semantics

I am fluent in over six million forms of communication
Distributional semantics

“You shall know a word by the company it keeps”
Firth (1957: 11)

Semantic knowledge → knowing *when* to use words
Contexts of use are a source of semantic information

| that. He was stood in front of me in the queue the other day and [unclear]. On the station he bought a and a cup of tea. He was surprised be located, how to prepare a salami, and what to do if you should come was quite expensive so I've bought a in the shop instead. That's a normal probably use to describe an indifferent . ‘A bit too smooth, though.’ ‘He nowhere till I've had a hot pastrami.’ We crowded into a mêlée like the that knowing how to make a Marmite would be enough. I pressed on. They for a stroll to the pub for a drink and a , they had spent nearly seventeen but I weren't sure if it was my fish paste or not! Shit! Just got a whiff as soon fat-free yoghurt. Supper Wholemeal with low-fat cream cheese and banana and if not, whether he should get a in a pub instead, and if so, whether of there. Well I like to have a toasted for dinner. I forget about it. Yeah, but up a [pause] plate Mhm. and I took the over Mhm. and I eat it and I went,
<table>
<thead>
<tr>
<th>that. He was stood in front of me in the sandwich queue the other day and [unclear] .</th>
</tr>
</thead>
<tbody>
<tr>
<td>On the station he bought a sandwich and a cup of tea. He was surprised be located, how to prepare a salami sandwich, and what to do if you should come.</td>
</tr>
<tr>
<td>was quite expensive so I've bought a sandwich in the shop instead. That's a normal sandwich. ‘A bit too smooth, though.’ ‘He</td>
</tr>
<tr>
<td>probably use to describe an indifferent sandwich. ’ We crowded into a mêlée like the</td>
</tr>
<tr>
<td>nowhere till I've had a hot pastrami sandwich. ’ We crowded into a mêlée like the</td>
</tr>
<tr>
<td>that knowing how to make a Marmite sandwich would be enough. I pressed on. The</td>
</tr>
<tr>
<td>for a stroll to the pub for a drink and a sandwich, they had spent nearly seventeen</td>
</tr>
<tr>
<td>but I weren't sure if it was my fish paste sandwich or not! Shit! Just got a whiff as soon</td>
</tr>
<tr>
<td>fat-free yoghurt. Supper Wholemeal sandwich with low-fat cream cheese and banana</td>
</tr>
<tr>
<td>and if not, whether he should get a sandwich in a pub instead, and if so, whether</td>
</tr>
<tr>
<td>of there. Well I like to have a toasted sandwich for dinner. I forget about it. Yeah, but</td>
</tr>
</tbody>
</table>
| up a [pause] plate Mhm. and I took the sandwich over Mhm. and I eat it and I went,
then [unclear] It was really part of the Mhm. Mhm. So did he [unclear] Do together.’ Hastings knew he'd got the Mhm. on Sunday night, and while he to be generally accepted that their x is by no means sinecure. Accordingly, the Hawick-based knitters showed x opportunities for at least 50 skilled poet John Wain was clearly doing his x . One aspect of the Lewis regime Write-in: I could do a better x if I knew more about Line he was sacked from the manager's x at Preston in 1981 he immediately told Arena today: ‘I go out and do a x on anyone who is giving our top to be ‘professionalized’, experts at our x . But sadly our world suffers because in the structure clearly identified by x descriptions and departmental As is the norm in such projects, every x turned out twice as extensive and grinning from ear to ear with his latest x . He has landed a plum role as the courses (part teaching, part practical x experience), while universities tend to
<table>
<thead>
<tr>
<th>Job</th>
</tr>
</thead>
</table>
| then [unclear] It was really part of the job. Mhm. Mhm. So did he [unclear] Do together.’ Hastings knew he'd got the job on Sunday night, and while he to be generally accepted that their job is by no means sinecure. Accordingly, the Hawick-based knitters showed job opportunities for at least 50 skilled poet John Wain was clearly doing his job. One aspect of the Lewis regime Write-in: I could do a better job if I knew more about Line he was sacked from the manager's job at Preston in 1981 he immediately told Arena today: ‘I go out and do a job on anyone who is giving our top to be ‘professionalized’, experts at our job in the structure clearly identified by job descriptions and departmental As is the norm in such projects, every job turned out twice as extensive and grinning from ear to ear with his latest job. He has landed a plum role as the courses (part teaching, part practical job experience), while universities tend to
of government. We will give a Cabinet responsibility for the Citizen's
the notes on clauses. I hope that the will clear that up when he replies.
rapidly-declining stocks. Fisheries John Crosbie denied, however, that
was elected as LDP leader and Prime in August 1989 [see pp. 36849-50].
party. My right hon. Friend the Prime was absolutely right to describe it as
in May or June [see ED 67]. Fisheries Jan Henry Olsen said a quota would
Dame Cath Tizard. Prime: Jim Bolger (since October 1990;
initiative on Aug. 15 the Iranian Foreign, Ali Akbar Velayati, described it as
The new Science and Technology sees information as an instrument of
Majorism isn't working? The Prime as the right hon. Gentleman is now
it's a moral problem, problem. The said, no it isn't, it's an economic
civil servant Sir Humphrey would tell his
civil servant Sir Humphrey would tell his
trade union paper Hodolmor, the new of Labour, Choyjamtsyn
of government. We will give a Cabinet minister responsibility for the Citizen's notes on clauses. I hope that the minister will clear that up when he replies.

rapidly-declining stocks. Fisheries minister John Crosbie denied, however, that was elected as LDP leader and Prime Minister in August 1989 [see pp. 36849-50].

party. My right hon. Friend the Prime Minister was absolutely right to describe it as in May or June [see ED 67]. Fisheries Minister Jan Henry Olsen said a quota would

Dame Cath Tizard. Prime Minister : Jim Bolger (since October 1990; initiative on Aug. 15 the Iranian Foreign minister, Ali Akbar Velayati, described it as

The new Science and Technology minister sees information as an instrument of Majorism isn't working? The Prime Minister as the right hon. Gentleman is now

it's a moral problem, problem. The minister said, no it isn't, it's an economic civil servant Sir Humphrey would tell his minister whenever the hapless Hacker
trade union paper Hodolmor, the new minister of Labour, Choyjamtsyn
“[I]f we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution.”

Harris (1954: 156)
Example: *drink* and *sip*

Sentences from the COCA corpus:

- the pizzeria for a while, **drinking** a beer at a table
- hell, I'd meet you, **drink** a glass of beer or
- books. She changed her dress, **drank** a glass of cold water
- Willie picks up his cup, **drinks** some coffee, and leaves with
- men picked up their beers, **sipped** them, and put them back
- to trust his intuition. She **sipped** from the champagne glass and
- food itself. Even when he **sipped** his cold beer, it was
- Emily was no different. Kate **sipped** from her water bottle, then
Example: *drink* and *sip*

the pizzeria for a while, **drinking** a **beer** at a table

hell, I'd meet you, **drink** a glass of **beer** or

books. She changed her dress, **drank** a glass of cold **water**

Willie picks up his cup, **drinks** some **coffee**, and leaves with

men picked up their **beers**, **sipped** them, and put them back
to trust his intuition. She **sipped** from the **champagne** glass and

food itself. Even when he **sipped** his cold **beer**, it was

Emily was no different. Kate **sipped** from her **water** bottle, then

**Beverages**
Example: *drink* and *sip*

the pizzeria for a while, **drinking** a **beer** at a table
   hell, I'd meet you, **drink** a **glass** of **beer** or
books. She changed her dress, **drank** a **glass** of cold **water**
   Willie picks up his **cup**, **drinks** some **coffee**, and leaves with
   men picked up their **beers**, **sipped** them, and put them back
   to trust his intuition. She **sipped** from the **champagne glass** and
   food itself. Even when he **sipped** his cold **beer**, it was
   Emily was no different. Kate **sipped** from her **water bottle**, then

**Beverages**

**Containers for beverages**
Example: *drink* and *sip*

the **pizzeria** for a while, *drinking* a **beer** at a **table**

  hell, I'd meet you, *drink* a **glass** of **beer** or

books. She changed her dress, *drank* a **glass** of cold **water**

  Willie picks up his **cup**, *drinks* some **coffee**, and leaves with

  men picked up their **beers**, *sipped* them, and put them back

  to trust his intuition. She *sipped* from the **champagne glass** and

  **food** itself. Even when he *sipped* his cold **beer**, it was

Emily was no different. Kate *sipped* from her **water bottle**, then

**Beverages**

**Containers for beverages**

**Drinking and dining**
‘Bag-of-words’ approach

Based on the frequency of co-occurrence between words in a large corpus

Count how many times each word occurs with each other word within a set context window

E.g., collocates of the verbs *answer, carry, push, reply, and tell* within a +/- 2 word window in the COHA corpus (400 MW)

|       | question | lift | heavy | softly | ...
|-------|----------|------|-------|--------|-----
| answer| 5854     | 44   | 13    | 119    | ... |
| carry | 56       | 66   | 512   | 27     | ... |
| push  | 41       | 28   | 58    | 27     | ... |
| reply | 201      | 40   | 3     | 66     | ... |
| tell  | 229      | 16   | 36    | 81     | ... |
‘Bag-of-words’ approach

Co-occurrence counts often replaced by association scores
I.e., how strong is the association between two words, given the individual frequency of these words?

Typical association measure: Positive Pointwise Mutual Information (PPMI)

|       | question | lift   | heavy  | softly | ...
|-------|----------|--------|--------|--------|--------
| answer| 3.8523   | 1.0399 | 0      | 1.1807 | ...    |
| carry | 0        | 1.1074 | 2.21   | 0      | ...    |
| push  | 0        | 1.3181 | 1.1003 | 0.4276 | ...    |
| reply | 0.7709   | 1.2347 | 0      | 0.8814 | ...    |
| tell  | 0        | 0      | 0      | 0      | ...    |
‘Bag-of-words’ approach

The rows of the matrix are called vectors

→ vector space models

<table>
<thead>
<tr>
<th></th>
<th>question</th>
<th>lift</th>
<th>heavy</th>
<th>softly</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>answer</td>
<td>3.8523</td>
<td>1.0399</td>
<td>0</td>
<td>1.1807</td>
<td>...</td>
</tr>
<tr>
<td>carry</td>
<td>0</td>
<td>1.1074</td>
<td>2.21</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>push</td>
<td>0</td>
<td>1.3181</td>
<td>1.1003</td>
<td>0.4276</td>
<td>...</td>
</tr>
<tr>
<td>reply</td>
<td>0.7709</td>
<td>1.2347</td>
<td>0</td>
<td>0.8814</td>
<td>...</td>
</tr>
<tr>
<td>tell</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The matrix is often reduced to a lower number of dimensions (e.g., by means of Singular Value Decomposition)
**‘Bag-of-words’** approach

Abstract distributional-semantic features corresponding to a large set of collocates

Vectors with similar values are expected to correspond to words with similar meaning

<table>
<thead>
<tr>
<th></th>
<th>(column 1)</th>
<th>(column 2)</th>
<th>(column 3)</th>
<th>(column 300)</th>
</tr>
</thead>
<tbody>
<tr>
<td>answer</td>
<td>11.662463</td>
<td>2.00896724</td>
<td>8.810539</td>
<td>-0.2389049</td>
</tr>
<tr>
<td>carry</td>
<td>21.827765</td>
<td>4.71476816</td>
<td>-11.974389</td>
<td>-0.52263</td>
</tr>
<tr>
<td>push</td>
<td>22.095771</td>
<td>13.130336</td>
<td>-6.027978</td>
<td>0.8539545</td>
</tr>
<tr>
<td>reply</td>
<td>15.407709</td>
<td>1.90698674</td>
<td>13.22548</td>
<td>-0.246191</td>
</tr>
<tr>
<td>tell</td>
<td>7.926409</td>
<td>0.06556502</td>
<td>4.79983</td>
<td>-0.3177306</td>
</tr>
</tbody>
</table>
Similarity

Semantic similarity is measured by mathematical similarity between word vectors

Most common measure: cosine

1: the vectors are identical

0: maximally dissimilar

<table>
<thead>
<tr>
<th></th>
<th>answer</th>
<th>carry</th>
<th>push</th>
<th>reply</th>
<th>tell</th>
</tr>
</thead>
<tbody>
<tr>
<td>answer</td>
<td>1</td>
<td>0.1871</td>
<td>0.2960</td>
<td>0.9241</td>
<td>0.6461</td>
</tr>
<tr>
<td>carry</td>
<td>0.1871</td>
<td>1</td>
<td>0.5787</td>
<td>0.1622</td>
<td>0.1514</td>
</tr>
<tr>
<td>push</td>
<td>0.2960</td>
<td>0.5787</td>
<td>1</td>
<td>0.2581</td>
<td>0.2314</td>
</tr>
<tr>
<td>reply</td>
<td>0.9241</td>
<td>0.1622</td>
<td>0.2581</td>
<td>1</td>
<td>0.6774</td>
</tr>
<tr>
<td>tell</td>
<td>0.6461</td>
<td>0.1514</td>
<td>0.2314</td>
<td>0.6774</td>
<td>1</td>
</tr>
</tbody>
</table>
Benefits

- Data-driven: more objective than ‘intuitive’ approach
- No manual intervention needed
- No limits on the number of lexical items
- Precise quantification
- Robust, adequately reflects semantic intuitions
  - Correlates with human performance in various tasks (e.g., Landauer et al. 1998, Lund et al. 1995)
  - Evidence for psychological adequacy (Andrews & Vigliocco 2008)

Using distributional semantics

- Distributional semantics is a robust way to capture semantic similarity, widely used in NLP

- How can it be used in linguistic research? Two methods:
  - Distributional semantic plots
    To visualize the semantic spread of a set of words
  - Distributional clustering
    To partition semantic development into stages

- Case studies in historical linguistics
Productivity

- The range of lexical items that can be used in the slots of a construction
- E.g., verbs in the “hell-construction”: V the hell out of NP (Perek 2014, 2016)

  You scared the hell out of me!

  I enjoyed the hell out of that show!

  But you drove the hell out of it!

  I've been listening the hell out of your tape.

  I voiced the hell out of ‘b’ (heard at GURT 2014, Georgetown)


The *hell*-construction in the COHA

- Recent construction: first instances in the 1930s
- Increasingly popular
- More and more verbs in the construction
- But how different are these verbs?
Distributional semantic plots

- Method to visualise the semantic space filled by a certain set of words
- Pairwise semantic distances are derived from a distributional semantic model
- Converted to a set of coordinates and plotted
  - E.g., with multidimensional scaling (MDS) or t-SNE (Van der Maaten & Hinton 2008)
  - Place objects in a 2-dimensional space such that the between-object distances are preserved as well as possible

Red: emotions, feelings, thoughts, mental activities
Blue: violent contact, exertion of force
Two domains of predilection

- Cognition verbs
  
  bother, disappoint, shock, startle, worry
  adore, enjoy, impress, love, want
  analyze, explain, understand

- Verbs of hitting and other forceful actions
  
  beat, knock, hit, kick, slap
  push, squeeze, twist
  blast, kill, shoot
The *way*-construction

- **Verb one’s way PP** (Perek 2016)
  
  *We pushed our way into the pub.*

- Focus on the “path-creation” use: the verb refers to the means what enables motion of the subject
  
  *They hacked their way through the jungle.*

- Vs. “manner” or “incidental-action”
  
  *They trudged their way through the snow.*
  
  *He whistled his way across the room.*

---

Data

- Relatively stable in frequency
- More and more verbs are used in the construction
Clear concrete/abstract divide in the distributional semantic plot

Higher density of verbs describing forceful actions (cut, push, kick, ..), especially in earlier periods.
From period 2 onwards: ingestion (*eat, drink, nibble, puff, sip, smoke, ..*), commerce & finance (*buy, export, fund, invest, pay, spend, ..*), misconduct (*bribe, bully, cheat, conspire, kill, murder, plot, rape, trick, ..*)
From period 3 onwards: social interaction (chat, chatter, joke, kid, nod, quarrel, talk), emotion (grin, laugh, smile, shrug, laugh), cognition (brood, fret, puzzle, think, worry)
The path-creation sense

- Many new verb classes refer to unusual ways to cause motion: interaction, commerce, cognition, etc.

- These new uses involve abstract, metaphorical motion:
  
  \[T\]hey talk about Uncle Paul having bought his way into the Senate!

  I sit and watch […], grazing my way through a muffuletta.

- Main semantic development: the construction becomes more and more open to encoding abstract motion
Periodization

- Distributional semantic plots are a useful tool to observe the development of constructions.
- However, it is limited by the arbitrary division of the data.
  - Periods of same length
  - Might not be consistent with regards to semantics
- Changes are assessed impressionistically rather than inferred quantitatively.
- This relates to the problem of periodization: how to reliably identify stages of change in the data?
Periodization

- Variant of agglomerative clustering algorithm
  - Periods are grouped according to their similarity, following some pre-defined criteria
  - Only time-adjacent periods can be merged

Distributional clustering

- VNC on the basis of the meaning of words attested in a construction at different points in time (Perek & Hilpert 2017)

Proposal:
- Use distributional semantics to build representations of the semantic range of a construction
- Submit these representations to VNC

Period vectors

- For each period, extract the semantic vector of each verb in the distribution of the construction.

- Add all vectors and divide by the number of verbs: this is the period vector.

  \[
  \begin{array}{cccc}
  \text{make} & 14.09814 & -4.231832 & -1.844898 \ldots & 0.06963598 \\
  \text{find} & 15.59443 & -2.022215 & 0.561186 \ldots & -0.5778517 \\
  \text{push} & 22.09577 & 13.130336 & -6.027978 \ldots & 0.8539545 \\
  \text{Sum} & 51.78834 & 6.876289 & -7.311691 \ldots & 0.3457388 \\
  \hline
  /3 & 17.26278 & 2.292096 & -2.43723 \ldots & 0.1152463 \\
  \end{array}
  \]

- “Semantic average” of the distribution.

- Features of the period vector reflect semantic properties of the verbs attested in the period.
The distributional clustering algorithm

- Starting point: data partitioned into “natural” time periods (years, decades, etc.)

1. Measure the similarity between the period vectors of all pairs of adjacent periods (e.g., 1830s-1840s, 1840s-1850s, etc.).
2. Merge the two periods that are the most similar.
3. Calculate the period vector of the merger as the mean between the vectors of its constituent periods.
4. Repeat until all periods have been merged.
The *hell*-construction

VNC dendrogram

Decades

Summed cosine distance

The path-creation way-construction
Interim summary

- The shapes of the dendrograms indicate different historical scenarios:
  - *Hell*-construction: gradually expanding construction
  - *Way*-construction: variations in distribution

- How to characterize each period?
  - The distributional-semantic features are highly abstract and not directly interpretable
  - The only way to interpret semantic changes is to look at the verbs themselves
Interpreting the dendrograms

- **1830s – 1870s**
  
  *hew, shape, explore, carve, track, enforce, shoulder, etc.*
  
  Concrete, physical actions, literal creation of a path

- **1890s – 2000s**
  
  *joke, bellow, chatter, snarl, spit, laugh, talk, bully, etc.*
  
  More abstract: communication, social interaction, etc.

- **1880s: transition period**
  
  *guess, buy, smell, stammer, beg, think, pay, etc.*
  
  *bore, pierce, feel, wear, melt, trace, burn, etc.*

- **Gradual change from mostly concrete to more abstract verbs, in line with previous findings**
Summary

- Distributional period clustering provides precise quantitative measurement to impressionistic observations.
- Models different kinds of change with dendrograms.
- Results are in line with semantic plots, but the timing of changes is measured more objectively.
Conclusion

- Distributional semantics is a promising tool for studies on productivity (and more)
- Turns the informal notion of meaning into a quantified representation
- Gives a semantic interpretation to changes in productivity
Theory?

- Such methods can inform theories of language change
- For instance, in diachronic construction grammar (Traugott & Trousdale 2013)
  - Grammar seen as inventory of form-meaning pairs, related in a taxonomic hierarchy (Goldberg 1995)
  - In diachrony: creation of new constructions, changes in existing ones, change in relations between constructions
- The *hell*-construction becomes more productive
- The *way*-construction becomes more productive *and* more schematic

Thanks for your attention!

f.b.perek@bham.ac.uk
www.fperek.net