Bloomfield’s curse

Bloomfield (1933: 274): “The lexicon is really an appendix of the grammar, a list of basic irregularities”

So . . . nothing to be seen here?
Bloomfield’s curse

Bloomfield (1933: 274): “The lexicon is really an appendix of the grammar, a list of basic irregularities”

- So ... nothing to be seen here?
- Or: more organizational principles than often thought?
Bloomfield’s curse

Bloomfield (1933: 274): “The lexicon is really an appendix of the grammar, a list of basic irregularities”

- So . . . nothing to be seen here?
- Or: more organizational principles than often thought?
- Our perspective: There is a universal continuous conceptual similarity structure between entities (objects, events, . . .)
Bloomfield’s curse

Bloomfield (1933: 274): “The lexicon is really an appendix of the grammar, a list of basic irregularities”

- So . . . nothing to be seen here?
- Or: more organizational principles than often thought?
- Our perspective: There is a universal continuous conceptual similarity structure between entities (objects, events, . . .)
- These spaces bias (1) learnability, (2) transferability, and even (3) adult categorical structure and hence shape lexica
Asymmetric overextension errors

- **Our focus:** learnability and developmental pathways of lexical domains
Asymmetric overextension errors

- Our focus: learnability and developmental pathways of lexical domains
- Revealing cases: asymmetric overextension errors
  - General case: $a$ means $A$ and $b$ means $B$
  - Children use $a$ to refer to $B$, but not $b$ to refer to $A$

Some known cases:
- Dutch: *op* 'horizontal, stable support' for *tenuous support*, but not *aan* 'tenuous support' for 'horizontal stable support'
- English: *blue* for 'purple', but not *purple* for 'blue'
- Dutch: *leggen* 'lay' for 'put, set', but not *zetten* 'put, set' for 'lay'

3/19
Asymmetric overextension errors

- Our focus: learnability and developmental pathways of lexical domains
- Revealing cases: asymmetric overextension errors
  - General case: \(a\) means \(A\) and \(b\) means \(B\)
  - Children use \(a\) to refer to \(B\), but not \(b\) to refer to \(A\)
- Some known cases
  - Dutch: \textit{op} ‘horizontal, stable support’ for ‘tenuous support’, but not \textit{aan} ‘tenuous support’ for ‘horizontal stable support’,
  - English: \textit{blue} for ‘purple’, but not \textit{purple} for ‘blue’
  - Dutch: \textit{leggen} ‘lay’ for ‘put, set’, but not \textit{zetten} ‘put, set’ but for ‘lay’
Our method: inferring maps from cross-linguistic data
General approach: Typology reveals conceptual space

- Gentner & Bowerman (2009): **Typological Prevalence Hypothesis**
  - The more languages refer two entities with a single label, the more cognitively similar they are
  - The more similar a group of entities is, the easier it is to learn a category extending over them
General approach: Elicit data

Ask speakers of a sample of languages to describe a series of situations

English
General approach: Elicit data

Ask speakers of a sample of languages to describe a series of situations

Dutch

- **IN**: A bowl with an apple inside.
- **OP**: A desk.
- **OM**: A person with a doughnut on their finger.
- **AAN**: A picture on a wall with a bird.
General approach: Elicit data

Ask speakers of a sample of languages to describe a series of situations

Tiriyo

[Diagram showing the words 'TAO', 'TAE', and 'PĒKĒ']
<table>
<thead>
<tr>
<th>situation</th>
<th>English</th>
<th>Dutch</th>
<th>Tiriyo</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPLE IN BOWL</td>
<td>in</td>
<td>in</td>
<td>tao</td>
</tr>
<tr>
<td>PAINTING ON WALL</td>
<td>on</td>
<td>aan</td>
<td>pēkē</td>
</tr>
<tr>
<td>RING ON FINGER</td>
<td>on</td>
<td>om</td>
<td>tae</td>
</tr>
<tr>
<td>PENCIL ON TABLE</td>
<td>on</td>
<td>op</td>
<td>tae</td>
</tr>
</tbody>
</table>
Dutch in the PCA space (components 1 and 3)
Case #1: acquiring spatial prepositions in Dutch and English
Data

- Gentner & Bowerman (2009): naming spatial relations in Dutch and English
  - English children make next to no errors
  - Dutch children use in for in and op for op correctly
  - Dutch children overextend op to aan and om situations but hardly ever aan or om to op

English

Dutch
Data

- Gentner & Bowerman (2009): naming spatial relations in Dutch and English
  - English children make next to no errors
  - Dutch children use *in* for **IN** and *op* for **OP** correctly

English

- *IN*
- *ON*

Dutch

- *IN*
- *OP*
- *OM*
- *AAN*
Data

- Gentner & Bowerman (2009): naming spatial relations in Dutch and English
  - English children make next to no errors
  - Dutch children use *in* for **IN** and *op* for **OP** correctly
  - Dutch children overextend *op* to **AAN** and **OM** situations
  - But hardly ever *aan* or *om* to **OP**
Modelling category learning

- General idea of learning
Modelling category learning

- General idea of learning
- Model receives pairs of term and situation
  - Sampling term-situation pairs on basis of term distribution in CDS
Modelling category learning

- General idea of learning
- Model receives pairs of term and situation
  - Sampling term-situation pairs on basis of term distribution in CDS
- Incrementally integrates into knowledge base
Modelling category learning

- General idea of learning
- Model receives pairs of term and situation
  - Sampling term-situation pairs on basis of term distribution in CDS
- Incrementally integrates into knowledge base
- Various cognitive models give very similar results
Dutch in the PCA space (components 1 and 3)
Beekhuizen, Fazly & Stevenson (2014): Model simulates asymmetrical overextension errors due to lay-out space

Overextension patterns
Are English speakers insensitive to the ‘strange’ lay-out of their *on* category?
Are English speakers insensitive to the ‘strange’ lay-out of their on category?

Task: judge adequacy of description (e.g. the apple is in the bowl for ‘apple in bowl’)

Prediction: faster judgements for more prototypical situations
Adult category structure (work with Nick Lester, UCSB)

- Are English speakers insensitive to the ‘strange’ lay-out of their on category?
- Task: judge adequacy of description (e.g. the apple is in the bowl for ‘apple in bowl’)
- Prediction: faster judgements for more prototypical situations
- Preliminary results: distance to cross-linguistic prototype highly predictive of RT

![Graph showing component 1 vs. component 3 with data points labeled from 1 to 65. The graph includes markers for 'in', 'aan', 'op', and 'om'.]
Case #2: acquiring color terms in Russian and English (dissociating crosslinguistic bias from perception)
Data

- **English**: Bateman (1915) 6-12yos – 8 color chips
  - *blue* for PURPLE, not vice versa
Data

- **English**: Bateman (1915) 6-12yos – 8 color chips
  - *blue* for *purple*, not vice versa

- **Russian**: Davies et al. (1998) 3-6yos – 12 color chips
  - *sinij* ‘dark blue’ for *light blue*, not *goluboj* ‘light blue’ for *dark blue*
  - *sinij* ‘dark blue’ for *purple*, not *fioletovyj* ‘purple’ for *dark blue*
  - *krasnyj* ‘red’ for *pink*, not *rozovyj* ‘pink’ for *red*
  - ...
Data

- English: Bateman (1915) 6-12yos – 8 color chips
  - *blue* for PURPLE, not vice versa
- Russian: Davies et al. (1998) 3-6yos – 12 color chips
  - *sinij* ‘dark blue’ for LIGHT BLUE, not *goluboj* ‘light blue’ for DARK BLUE
  - *sinij* ‘dark blue’ for PURPLE, not *fioletovyj* ‘purple’ for DARK BLUE
  - *krasnyj* ‘red’ for PINK, not *rozovyj* ‘pink’ for RED
  - ...

- Same learning approach
- **Contrasting two approaches:** using *Lab* space (perceptual space of color) versus crosslinguistic space
Results: Beekhuizen & Stevenson (2015; 2016)

- compare model’s rankings of term given chip with observed rankings
- crosslinguistic space (CL) for English: poor fit
- but it is **complementary** to perceptual for English!

<table>
<thead>
<tr>
<th></th>
<th>Russian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>perceptual</td>
<td>.91</td>
<td>.96</td>
</tr>
<tr>
<td>CL</td>
<td>.91</td>
<td>.91</td>
</tr>
<tr>
<td>perceptual+CL</td>
<td>.90</td>
<td>.98</td>
</tr>
<tr>
<td>no-development baseline</td>
<td>.81</td>
<td>.95</td>
</tr>
</tbody>
</table>

Fit with child data (Kendall $\tau_b$ for term rankings).
Final thoughts

- Deriving semantic space from crosslinguistic data provides novel way of modeling semantic space
- Test case: asymmetric overextension errors
- Space goes beyond mere perception: case of COLOR
- Not just acquisition: organization of adult categories reflects this space too.
Thank you!