NooJ in the Humanities: Phrasal Verb Usage in the Works of British and American Authors

NooJ 2020 Virtual Conference

Peter A. Machonis
Florida International University
English phrasal verbs (PV) or verb-particle combinations

• *figure out, finish up, turn on*, etc.
• particle movement adds to the difficulty of accurately locating them
• *figure out the answer* ⇔ *figure the answer out*
• characteristic trait of the English language
• steady growth from the 13th to the 17th century, slight drop during in the 18th century, expansion in the 19th century
• often attributed to an American influence by grammarians
• “creativeness that surpasses anything else in our language”
• “index of emotional involvement of the speaker”
• “little evidence for the universal assumption that PV are more typical of American English.”
NooJ in the humanities (and linguistics)

- PV grammar
- PV electronic dictionary
- series of disambiguation grammars
- adverbial and adjectival expression filters
- idiom dictionaries
- GOAL: parse the complete works of Charles Dickens and his American counterpart, Herman Melville and compare their PV usage
# NooJ PV Dictionary

<table>
<thead>
<tr>
<th>Entry</th>
<th>S-lemma</th>
<th>Category</th>
<th>PLX</th>
<th>FR</th>
<th>Lexample</th>
</tr>
</thead>
<tbody>
<tr>
<td>beef</td>
<td>beef</td>
<td>V</td>
<td>ASK</td>
<td>renforcer</td>
<td>the proposal</td>
</tr>
<tr>
<td>bend</td>
<td>bend</td>
<td>V</td>
<td>BEND</td>
<td>tordre</td>
<td>the credit card</td>
</tr>
<tr>
<td>bind</td>
<td>bind</td>
<td>V</td>
<td>FIND</td>
<td>bander</td>
<td>the wound</td>
</tr>
<tr>
<td>block</td>
<td>block</td>
<td>V</td>
<td>ASK</td>
<td>obstruer</td>
<td>the sink</td>
</tr>
<tr>
<td>blow</td>
<td>blow</td>
<td>V</td>
<td>BLOW</td>
<td>agrandir</td>
<td>the photo</td>
</tr>
<tr>
<td>blow</td>
<td>blow</td>
<td>V</td>
<td>BLOW</td>
<td>exagérer</td>
<td>the scandal</td>
</tr>
<tr>
<td>blow</td>
<td>blow</td>
<td>V</td>
<td>BLOW</td>
<td>exploser</td>
<td>the building</td>
</tr>
<tr>
<td>blow</td>
<td>blow</td>
<td>V</td>
<td>BLOW</td>
<td>gentler</td>
<td>the balloons</td>
</tr>
<tr>
<td>boil</td>
<td>boil</td>
<td>V</td>
<td>ASK</td>
<td>bouillir</td>
<td>some water</td>
</tr>
</tbody>
</table>

**Part | synonym | Syntax | SynSem**
--- | ------- | ------ | -------
up    | strengthen | -     | PV+FXC+NG Hum+NO NHum+NL NHum
up    | bend completely | - | PV+FXC+NG Hum+NO NHum+NL NHum+NoPart Part
up    | put bandage on | - | PV+FXC+NG Hum+NL NHum+NL NHum+NoPart Part
up    | obstruct | - | PV+FXC+NG Hum+NO NHum+NL NHum+NoPart +NlV Part
up    | enlarge | - | PV+FXC+NG Hum+NO NHum+NL NHum
up    | exaggerate | - | PV+FXC+NG Hum+NO NHum+NL NHum+NL V Part
up    | explode | - | PV+FXC+NG Hum+NO NHum+NL NHum+NL NHum+NL V Part +NL V
up    | inflate | - | PV+FXC+NG Hum+NO NHum+NL NHum
up    | boil | - | PV+FXC+NG Hum+NL NHum+NoPart Part +NL V
NooJ PV Grammar
Automatic PV Recognition is more complex than for other multiword expressions: confusion with verbs followed by simple prepositions

- *Do you remember what I asked you in Rome?* (verb + prepositional phrase)
- *Did you ask the prince in when he arrived?* (PV)
- *had a strange smile on her thin lips* (verb + prepositional phrase)
- *had her hat and jacket on* (PV)
- the tears *rolled down* the poor child's face (verb + prepositional phrase)
- *rolling back* the table-cloth (PV)
- She *took* her feet *off* the fender, pushed back her chair (verb + prepositional phrase)
- He *took off* the cocked hat again (PV)
Not always easy to distinguish prepositional phrase from PV

- Mr. Sikes **pulled off** Oliver's cap and threw it into a corner
  - PV
- 'Pull that cap off your eyes, and hold up your head, sir.'
  - PV or PP ??????
- **Count out** the money for them
  - PV
- **Count up** to twelve
  - PV or PP ??????
Using NooJ for Automatic PV Recognition

<table>
<thead>
<tr>
<th>TEXT</th>
<th>Potential PV strings identified by NooJ</th>
<th>Correct PV (continuous)</th>
<th>Correct PV (discontinuous)</th>
<th>Incorrect PV (Prepositions)</th>
<th>Incorrect PV (Misidentifications)</th>
<th>Percentage of incorrect</th>
<th>PRECISION: Percentage of PV correctly identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portrait of a Lady (2010)</td>
<td>583</td>
<td>405</td>
<td>83</td>
<td>44</td>
<td>51</td>
<td>16.30%</td>
<td>83.70%</td>
</tr>
<tr>
<td>Portrait of a Lady (2016)</td>
<td>658</td>
<td>426</td>
<td>152</td>
<td>62</td>
<td>18</td>
<td>12.16%</td>
<td>87.84%</td>
</tr>
<tr>
<td>Portrait of a Lady (2018)</td>
<td>636</td>
<td>426</td>
<td>152</td>
<td>55</td>
<td>3</td>
<td>9.12%</td>
<td>90.88%</td>
</tr>
<tr>
<td>Portrait of a Lady - six particles (2020)</td>
<td>522</td>
<td>393</td>
<td>124</td>
<td>2</td>
<td>3</td>
<td>0.96%</td>
<td>99.04%</td>
</tr>
</tbody>
</table>

⇒ 1881 novel *The Portrait of a Lady* (233,102 word forms)
⇒ *PRECISION: True Positives/(True Positives + False Positives)*
Hiltunen (1994)

- three levels of PV frequency
  - high
    - *out, up*
  - mid
    - *down, away*
  - low
    - *back, off*
Examples of discontinuous PV involving two or more word forms correctly identified by NooJ

- I still held her forcibly down; *(Great Expectations)*
- She pushed the candle impatiently away; *(Oliver Twist)*
- The sea had jeeringly kept his finite body up, *(Moby Dick)*
- and wear his daily life out grain by grain *(The Mystery of Edwin Drood)*
- Shall I put any of those little things up with mine? *(The Pickwick Papers)*
To linen that you want sent to the washerwoman's? I know a delightful washerwoman outside, that comes for my things twice a week; and, by Jove!—how devilish lucky!—this is the day she calls. Shall I put any of those little things up with mine? Don't say anything about the trouble. Confound and rise! if one gentleman under a cloud is not to put himself a little out of the way to assist another gentleman in the same condition, what's human nature?
Example of noise not removed (Dickens)

- wearing a very large sword, and **boots up** to his hips,
- Jasper [ ... ] **turns** his **back** to mix a jug of mulled wine at the fire
- drawn **close up** before the brightest possible little fire
  - stepped **close up** to him on tiptoe [removed]
  - close up to, **PREP** but **step up** [silence]
- not a little indignant to **hear** a young urchin roaring **out** some jolly song
  - **roaring out** some jolly song [correctly identified]
- when he **heard** his adversary call **out** for a cessation of
Example of noise not removed (Melville)

- His hair **had** fallen **out**, and left him very bald
  - fallen,N+Nb=p
- I paused; then went **close up** to him;
- they say he "**takes** a walk **up** Ladder-lane
  - exclude **up** from locative disambiguation grammar
- We looked **clean up** through the unbearded timbers
- We **looked** clean **up** [silence]
  - NOT noted because **clean** is NOT an adverb
**RESULTS (Dickens)**

<table>
<thead>
<tr>
<th>DATE</th>
<th>TEXT</th>
<th>WORD FORMS</th>
<th>Phrasal Verbs (noise deleted)</th>
<th>Phrasal Verbs per 1,000 words of text</th>
<th>NooJ results</th>
<th>noise</th>
<th>PRECISION*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1836</td>
<td>The Pickwick Papers</td>
<td>308,843</td>
<td>1051</td>
<td>3.40</td>
<td>1068</td>
<td>17</td>
<td>98.4%</td>
</tr>
<tr>
<td>1837</td>
<td>Oliver Twist</td>
<td>161,518</td>
<td>636</td>
<td>3.94</td>
<td>648</td>
<td>12</td>
<td>98.1%</td>
</tr>
<tr>
<td>1843</td>
<td>A Christmas Carol</td>
<td>29,185</td>
<td>112</td>
<td>3.84</td>
<td>115</td>
<td>3</td>
<td>97.4%</td>
</tr>
<tr>
<td>1849</td>
<td>David Copperfield</td>
<td>363,485</td>
<td>1240</td>
<td>3.41</td>
<td>1265</td>
<td>25</td>
<td>98.0%</td>
</tr>
<tr>
<td>1859</td>
<td>A Tale of Two Cities</td>
<td>138,157</td>
<td>427</td>
<td>3.09</td>
<td>433</td>
<td>6</td>
<td>98.6%</td>
</tr>
<tr>
<td>1860</td>
<td>Great Expectations</td>
<td>188,948</td>
<td>794</td>
<td>4.20</td>
<td>807</td>
<td>13</td>
<td>98.4%</td>
</tr>
<tr>
<td>1870</td>
<td>The Mystery of Edwin Drood</td>
<td>97,844</td>
<td>301</td>
<td>3.08</td>
<td>302</td>
<td>1</td>
<td>99.7%</td>
</tr>
<tr>
<td>TOTALS and AVERAGE</td>
<td>1,287,980</td>
<td>4561</td>
<td>3.54</td>
<td>4638</td>
<td>77</td>
<td>98.34%</td>
<td></td>
</tr>
</tbody>
</table>

*PRECISION: True Positives/(True Positives + False Positives)
## RESULTS (Melville)

<table>
<thead>
<tr>
<th>DATE</th>
<th>TEXT</th>
<th>WORD FORMS</th>
<th>Phrasal Verbs (noise deleted)</th>
<th>Phrasal Verbs per 1,000 words of text</th>
<th>NooJ results</th>
<th>noise</th>
<th>PRECISION*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1846</td>
<td>Typee: A Romance of the South Sea</td>
<td>114,984</td>
<td>243</td>
<td>2.11</td>
<td>246</td>
<td>3</td>
<td>98.8%</td>
</tr>
<tr>
<td>1847</td>
<td>Omoo: Adventures in the South Seas</td>
<td>103,665</td>
<td>336</td>
<td>3.24</td>
<td>339</td>
<td>3</td>
<td>99.1%</td>
</tr>
<tr>
<td>1849</td>
<td>Redburn. His First Voyage</td>
<td>120,504</td>
<td>390</td>
<td>3.24</td>
<td>396</td>
<td>6</td>
<td>98.5%</td>
</tr>
<tr>
<td>1850</td>
<td>White Jacket or, the World on a Man-of-War</td>
<td>145,598</td>
<td>382</td>
<td>2.62</td>
<td>384</td>
<td>2</td>
<td>99.5%</td>
</tr>
<tr>
<td>1851</td>
<td>Moby Dick</td>
<td>218,390</td>
<td>574</td>
<td>2.63</td>
<td>579</td>
<td>5</td>
<td>99.1%</td>
</tr>
<tr>
<td>1853</td>
<td>Bartleby, The Scrivener A Story of Wall-Street</td>
<td>14,650</td>
<td>34</td>
<td>2.32</td>
<td>38</td>
<td>4</td>
<td>89.5%</td>
</tr>
<tr>
<td>1857</td>
<td>The Confidence-Man: His Masquerade</td>
<td>95,602</td>
<td>233</td>
<td>2.44</td>
<td>233</td>
<td>0</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>TOTALS and AVERAGE</td>
<td>813,393</td>
<td>2192</td>
<td>2.69</td>
<td>2215</td>
<td>23</td>
<td>98.96%</td>
</tr>
</tbody>
</table>

*PRECISION: True Positives/(True Positives + False Positives)
Future Work

- *away* is currently underrepresented as a mid-level frequency particle
  - *frighten away, wear away*
  - Sydney La Valley (2020)
- enhance the representation of the two low-level frequency particles *back* and *off*
  - *brush down, burn down, carry off, dry off*
- American: Herman Melville, James Fenimore Cooper, Washington Irving, Nathaniel Hawthorne, Harriet Beecher Stowe, Mark Twain, Edith Wharton
- British: Charles Dickens, Jane Austen, Walter Scott, the Bronte sisters, George Eliot, Thomas Hardy, Oscar Wilde
- NooJ: connecting the past to the present and the future of research
- Thank you for your attention