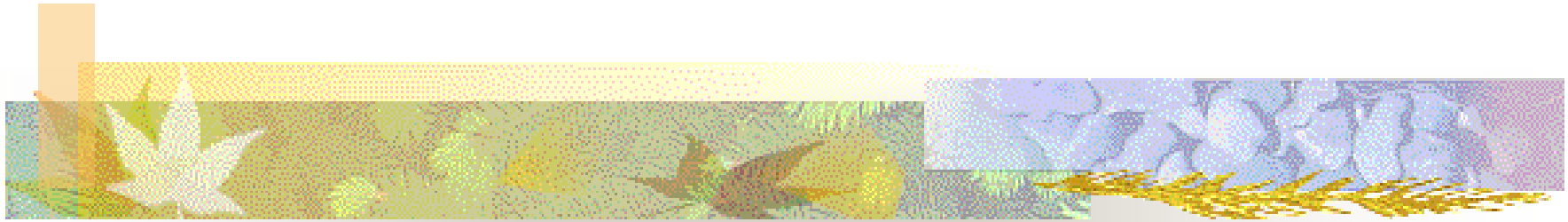


Symposium on Corpus Approaches to Lexicogrammar



LxGr2017

June 10, 2017

Edge Hill University

A Corpus-based Analysis of the Works of Dickens and Melville



Are Phrasal Verbs Typically American?

Peter A. Machonis

Florida International University



Kennedy (1920)

- English Phrasal Verbs (PV) = pleonastic or colloquial variants of simple verbs
 - *finish up* vs. *finish*
 - *cough up* vs. *pay*
- Konishi (1958): progressive increase from 13th to 17th century
- decrease in frequency in the 18th
 - prescriptive attitude?
- new expansion in the 19th century



The New Fowler's (2000:594)

- “Frequent in American English, it is clear that the use of PV began to increase in a noticeable manner in America from the early 19th century onward. From there, many have made their way to Britain during the 20th century, to widespread expressions of regret and alarm.”




Are Phrasal Verbs Typically American?

- Martin (1990): no proof of a great discrepancy between American and British usage
- Rohdenburg (2009): certain PV **are** typically American
- Hampe (2002): compositional PV = “index of emotional involvement” or “more specific in meaning than their bare-verb counterparts”



Thim (2012)

- “little evidence for the universal assumption that PV are more typical of American English”
- “the little attention Late Modern English – in particular the 19th century – has received”
- “Most of the 19th century is not covered at all”



Natural Language Processing & Historical Linguistics

- Automatically annotate PV, both continuous and discontinuous, in large corpora
 - That **turned around** the national economy
 - That **turned** the national economy **around**
 - That program **brought down** our crime rates
 - That program **brought** our crime rates **down**
- Novels of Charles Dickens & Herman Melville
 - Data on 19th century British and American usage



Using NooJ to count PV

<http://www.nooj4nlp.net/>

- Open source linguistic development environment
- Parses texts using large-scale dictionaries and grammars
- Instead of a part of speech tagger, a **Text Annotation Structure (TAS)** holds unsolved ambiguities
- Can process **discontinuous linguistic units**



Lexicon-Grammar

- Gross (1994, 1996): reproducibility of linguistic data in the form of exhaustive syntactic tables
- Manually constructed, containing both lexical and syntactic information
 - *up* (700 entries) Transitive and Neutral (e.g., *blow up*)
 - *out* (200 entries)
 - Other particles *away, back, down, in, off, over* (300 entries)

$N_0 :: N_{hum}$	$N_0 :: N-hum$	Verb	Particle	Example of N_1	$N_1 :: N_{hum}$	$N_1 :: N-hum$	$N_0 V N_1$	$N_1 V Part$	$N_1 V$	Synonym
+	+	beam	up	the aliens	+	+	-	+	-	transport by energy
+	+	bear	up	the weight	+	+	+	-	-	support
+	+	beat	up	the door	-	+	-	-	-	damage
+	+	beat	up	the eggs	-	+	+	-	-	beat
+	-	beat	up	the child	+	-	+	-	-	attack physically & hurt
+	+	beef	up	the proposal	-	+	-	-	-	strengthen
+	+	bend	up	the credit card	-	+	+	-	-	bend completely
+	-	bind	up	the wound	+	+	+	-	-	put bandage on
+	+	block	up	the sink	-	+	+	+	-	obstruct
+	+	blow	up	the balloons	-	+	-	-	-	inflate
+	+	blow	up	the building	+	+	-	+	+	explode
+	+	blow	up	the photo	-	+	-	-	-	enlarge
+	+	blow	up	the scandal	-	+	-	+	-	exaggerate
+	-	boil	up	some water	-	+	+	-	+	boil
+	+	bolster	up	Max	+	+	+	-	-	give hope to
+	+	bolster	up	the theory	-	+	-	-	-	support
+	+	boot	up	the computer	-	+	+	+	+	start

$N_0 =: N_{hum}$	$N_0 =: N_{-hum}$	Verb	Particle	Example of N_1	$N_1 =: N_{hum}$	$N_1 =: N_{-hum}$	$N_0 V N_1$	$N_1 V$ Part	$N_1 V$	Synonym
+	+	beam	up	the aliens	+	+	-	+	-	transport by energy
+	+	bear	up	the weight	+	+	+	-	-	support
+	+	beat	up	the door	-	+	-	-	-	damage
+	+	beat	up	the eggs	-	+	+	-	-	beat
+	-	beat	up	the child	+	-	+	-	-	attack physically & hurt
+	+	beef	up	the proposal	-	+	-	-	-	strengthen
+	+	bend	up	the credit card	-	+	+	-	-	bend completely
+	-	bind	up	the wound	+	+	+	-	-	put bandage on
+	+	block	up	the sink	-	+	+	+	-	obstruct
+	+	blow	up	the balloons	-	+	-	-	-	inflate
+	+	blow	up	the building	+	+	-	+	+	explode
+	+	blow	up	the photo	-	+	-	-	-	enlarge
+	+	blow	up	the scandal	-	+	-	+	-	exaggerate
+	-	boil	up	some water	-	+	+	-	+	boil
+	+	bolster	up	Max	+	+	+	-	-	give hope to
+	+	bolster	up	the theory	-	+	-	-	-	support
+	+	boot	up	the computer	-	+	+	+	+	start

NooJ

File Edit Lab Project Windows Info DICTIONARY

C:\Users\Peter\Documents\NooJ\en\Lexical Analysis\SIX Particles.dic
 Dictionary contains 1148 entries

Entry	S-L...	C.	FLX	FR	Nlexample	Part	synonym	SynSem
beef	beef	V	ASK	renforcer	the proposal	up	strengthen	PV+FXC+NOHum+NONHum+N1NHum
bend	bend	V	BEND	tordre	the credit card	up	bend comple...	PV+FXC+NOHum+NONHum+N1NHum+NoPart
bind	bind	V	FIND	bander	the wound	up	put bandage on	PV+FXC+NOHum+N1Hum+N1NHum+NoPart
block	block	V	ASK	obstruer	the sink	up	obstruct	PV+FXC+NOHum+NONHum+N1NHum+NoPart+N1VPart
blow	blow	V	BLOW	agrandir	the photo	up	enlarge	PV+FXC+NOHum+NONHum+N1NHum
blow	blow	V	BLOW	exagérer	the scandal	up	exaggerate	PV+FXC+NOHum+NONHum+N1NHum+N1VPart
blow	blow	V	BLOW	explorer	the building	up	explode	PV+FXC+NOHum+NONHum+N1Hum+N1NHum+N1VPart+N1V
blow	blow	V	BLOW	gonfler	the balloons	up	inflate	PV+FXC+NOHum+NONHum+N1NHum
boil	boil	V	ASK	bouillir	some water	up	boil	PV+FXC+NOHum+N1NHum+NoPart+N1V

Cancel

NooJ File Edit Lab Project Windows Info DICTIONARY

Dictionary contains 1256 entries

Entry	S-Lemma	Category	FLX	FR	Nlexample
beef	beef	V	ASK	renforcer	the proposal
bend	bend	V	BEND	tordre	the credit card
bind	bind	V	FIND	bander	the wound
block	block	V	ASK	obstruer	the sink
blow	blow	V	BLOW	agrandir	the photo
blow	blow	V	BLOW	exagérer	the scandal
blow	blow	V	BLOW	explorer	the building
blow	blow	V	BLOW	gonfler	the balloons
boil	boil	V	ASK	bouillir	some water

14 sec Cancel

Part	synonym	Syntax	SynSem
up	strengthen	-	PV+FXC+NOHum+NONHum+N1NHum
up	bend completely	-	PV+FXC+NOHum+NONHum+N1NHum+NoPartPart
up	put bandage on	-	PV+FXC+NOHum+N1Hum+N1NHum+NoPartPart
up	obstruct	-	PV+FXC+NOHum+NONHum+N1NHum+NoPart+N1VPart
up	enlarge	-	PV+FXC+NOHum+NONHum+N1NHum
up	exaggerate	-	PV+FXC+NOHum+NONHum+N1NHum+N1VPart
up	explode	-	PV+FXC+NOHum+NONHum+N1Hum+N1NHum+N1VPart+N1V
up	inflate	-	PV+FXC+NOHum+NONHum+N1NHum
up	boil	-	PV+FXC+NOHum+N1NHum+NoPartPart+N1V

14 sec Cancel



Fairly complex problem

- Possible discontinuous nature
 - *let out the dogs* ⇔ *let the dogs out*
- Confusion with verb followed by preposition
 - *Do you remember what I asked you in Rome?*
(preposition)
 - *Did you ask the prince in when he arrived?* (PV)
- Genuine ambiguity: resolvable from context
 - *Her neighbor was looking over the broken fence*



Fairly complex program

- Three disambiguation grammars that remove false PV (noise)
 - One looks at environment to the left of PV
 - Two examine environments to the right of PV
- Adverbial and adjectival expression filters
- Idiom dictionaries and grammars
- Machonis (2017) shows how NooJ removes inaccurate Text Annotation Structures (TAS)



NooJ identifies discontinuous PV

- I still **held** her forcibly **down**
- Joe **took** his hat **off** and stood
- If you **bring** the boy **back** with his head blown to bits by a musket,
- Canst thou not **drive** that old Adam **away**?
- The sea had jeeringly **kept** his finite body **up**, but drowned the infinite of his soul.
- he had that club-hammer there ... to **knock** some one's brains **out** with, I suppose,



Noise mainly comes from particles *in* and *on*

- *had a strange smile on her thin lips* ≠ PV
- *had her hat and jacket on* = PV
- reduce PV dictionary to only six particles
- 98% accuracy with only six particles
 - *out, up, down, away, back, off*



PV usage per 1,000 words of text

- Hiltunen (1994): six typical particles representing three levels of PV frequency
- high (*out, up*), mid (*down, away*), and low (*back, off*).
- Novels of Melville (1.3 million words)
- Novels of Dickens (4 million words)
- Project Gutenberg
- Yet still some noise (falsely annotated PV)



Difficulties counting PV

- Prepositional verbs
- Call on \neq PV dictionary
- Make up = PV dictionary
- Women make up half of the applicants \neq PV
- to make up some of our deficiencies = PV



Difficulties counting PV (cont.)

- Phrasal prepositional verbs
- The students **looked up to** the teacher ?
- When he was in Paris, he **looked Max up** (**up** Max) to see how he was doing. = PV
- he **looked up** to the flag, (*Typee*) = PV
- when I **looked up** to the warm, glowing, tropical sky = PV



NooJ identifies some intransitive PV

- Let's **get** the scaffolding **up**, to **get** the great umbrella **up**, etc.
- *get up* “rise” (intransitive)
- We **got up** early in the morning
- I **got up**, several times, and looked



Nouns mistaken for verbs

- They were a gloomy suite of rooms, in a lowering pile of **building up** a yard, (A Christmas Carol) \neq PV
- to keep the raw morning **air out**; (David Copperfield) \neq PV
- With this my aunt tied her **head up** in a handkerchief, \neq PV
- After that, she covered her **head up** with her apron again \neq PV



Nouns and prepositions mistaken for particles

- and **turned** his powerful **back** (Great Expectations) \neq PV
- But has she not **taken** me **down** stairs,
- in **keeping** all the provisions **up** stairs
- we saw him **taken up** the side and disappear
- **Running down** stairs, I quickly stated (Moby Dick)



Other noise, difficult to eliminate

- strange visions of outlandish things **does** the very name spirit **up** (*Typee*) \neq PV
- had been **done up** into a small bundle = PV
- and drew my head **close down** to hers \neq PV
- Don't straggle, my man. **Close up** here. ?
- I paused; then went **close up** to him; stooped over \neq PV (*Bartleby*)



Continuous vs. Discontinuous PV

- Most PV in our novels are continuous
- Discontinuous PV still frequent with one word form
- Dickens: 20% PV with one word form
- Melville: 12% PV with one word form
- Two word forms: 4 to 5%
- Three word forms: 1%
- *Away* and *up* most frequent with more than one word form

DATE	TEXT	WORD FORMS	Phrasal Verbs	Phrasal Verbs per 1,000 words of text
	CHARLES DICKENS			
1836	The Pickwick Papers	308,843	1068	3.46
1837	Oliver Twist	161,518	652	4.04
1838	Nicholas Nickleby	330,990	972	2.94
1840	The Old Curiosity Shop	221,855	693	3.12
1841	Barnaby Rudge	259,931	910	3.50
1843	A Christmas Carol	29,185	115	3.94
1843	Martin Chuzzlewit	345,529	1058	3.06
1846	Dombey and Son	363,526	1180	3.25
1849	David Copperfield	363,485	1265	3.48
1852	Bleak House	361,983	1215	3.36
1854	Hard Times	105,682	309	2.92
1855	Little Dorrit	344,932	1139	3.30
1859	A Tale of Two Cities	138,157	435	3.15
1860	Great Expectations	188,948	807	4.27
1864	Our Mutual Friend	333,783	1281	3.84
1870	The Mystery of Edwin Drood	97,844	302	3.09
	TOTALS and AVERAGE	3,956,191	13401	3.39

DATE	TEXT	WORD FORMS	Phrasal Verbs	Phrasal Verbs per 1,000 words of text
	HERMAN MELVILLE			
1846	Typee: A Romance of the South Sea	114,984	246	2.14
1847	Omoo: Adventures in the South Seas	103,665	344	3.32
1849	Mardi: And a Voyage Thither (vol. 1)	98,220	207	2.11
1849	Mardi: And a Voyage Thither (vol. 2)	103,555	228	2.20
1849	Redburn. His First Voyage	120,504	398	3.30
1850	White Jacket or, the World on a Man-of-War	145,598	385	2.64
1851	Moby Dick	218,390	579	2.65
1852	Pierre; or The Ambiguities	159,700	300	1.88
1853	Bartleby, The Scrivener A Story of Wall-Street	14,650	38	2.59
1855	Israel Potter	67,026	180	2.69
1856	The Piazza Tales	80,998	175	2.16
1857	The Confidence-Man: His Masquerade	95,602	233	2.44
	TOTALS and AVERAGE	1,322,892	3313	2.50

TEXT	WORD FORMS	Phrasal Verbs	Phrasal Verbs per 1,000 words of text
CHARLES DICKENS			
Complete novels and stories	3,956,191	13,401	3.39
HERMAN MELVILLE			
Complete novels and stories	1,322,892	3,313	2.50
Transcribed <i>Larry King Live</i> (January 2000)	228,950	541	2.36



Conclusion

- Thim (2012) “undercuts much received notion on the PV”
- New data on the history of PV, in particular to written usage in 19th century America and Britain
- Not necessarily an American phenomenon
- *Larry King Live* programs (January 2000)
- 2.36 PV per 1,000 words
- Perhaps not a purely spoken phenomenon either

Thank you for your attention



... Comments?

... Questions?

machonis@fiu.edu



References

- Bolinger, Dwight. 1971. *The Phrasal Verb in English*. Cambridge, MA: Harvard University Press.
- Brinton, Laurel J. 2013. Review of *Phrasal Verbs: The English Verb-Particle Construction and Its History*. By Stephan Thim. *Language* 89.3: 664-667.
- Burchfield, Robert William. 2000. *The New Fowler's Modern English Usage*. Rev. 3rd ed. Oxford; New York: Oxford University Press.
- Gross, Maurice. 1994. Constructing Lexicon-Grammars. *Computational Approaches to the Lexicon*, ed. by Atkins and Zampolli, 213-263. Oxford: Oxford University Press.
- _____. 1996. Lexicon Grammar. *Concise Encyclopedia of Syntactic Theories*, ed. by K. Brown and J. Miller, 244-258. New York: Elsevier.
- Hampe, Beate. 2002. Superlative Verbs: A corpus-based study of semantic redundancy in English verb-particle constructions. Tübingen: Gunter Narr Verlag.
- Hiltunen, Risto. 1994. On Phrasal Verbs in Early Modern English: Notes on Lexis and Style. *Studies in Early Modern English*, ed. by Dieter Kastovsky, 129-140. Berlin: Mouton de Gruyter.
- Kennedy, Arthur Garfield. 1920. *The Modern English Verb-adverb Combination*. Stanford: Stanford University Press.
- Konishi, Tomoshichi. 1958. The growth of the verb-adverb combination in English: A brief sketch. In Kazuo Araki, Taiichiro Egawa, Toshiko Oyama & Minoru Yasui (eds.), *Studies in English grammar and linguistics: A miscellany in honour of Takanobu Otsuka*, 117-128. Tokyo: Kenkyusha.

- 
- Machonis, Peter A.. 2010. English Phrasal Verbs: from Lexicon-Grammar to Natural Language Processing. *Southern Journal of Linguistics* 34.1: 21-48.
 - _____. 2012. *Sorting NooJ out to take Multiword Expressions into account*. *Automatic Processing of Various Levels of Linguistic Phenomena: Selected Papers from the NooJ 2011 International Conference*, ed. by Kristina Vučković, Božo Bekavac, and Max Silberztein, 152-165. Newcastle upon Tyne: Cambridge Scholars Publishing.
 - _____. 2017 (in press). Phrasal Verb Disambiguating Grammars: Cutting Out Noise Automatically. *Automatic Processing of Natural-Language Electronic Texts with NooJ*, ed. by Linda Barone, Max Silberztein, and Mario Monteleone . Cham, Switzerland: Springer International Publishing AG.
 - Martin, Pamela. 1990. *The Phrasal Verb : Diachronic Development in British and American English*. New York, NY: Columbia University Ed. D. dissertation.
 - NooJ: A Linguistic Development Environment. <http://www.nooj4nlp.net/>
 - Project Gutenberg. http://www.gutenberg.org/wiki/Main_Page
 - Silberztein, Max. 2016. *Formalizing Natural Languages: The NooJ Approach*. London: Wiley ISTE.
 - Thim, Stephan. 2012. *Phrasal Verbs: The English Verb-Particle Construction and Its History*. Berlin: Walter de Gruyter.
 - Rohdenburg, Günter. 2009. Grammatical Divergence between British and American English in the Nineteenth and Early Twentieth Centuries. *Current Issues in Late Modern English*, ed. by Ingrid Tieken-Boon van Ostade and Wim van der Wurff, 301-329. Bern: Lang.