



# **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

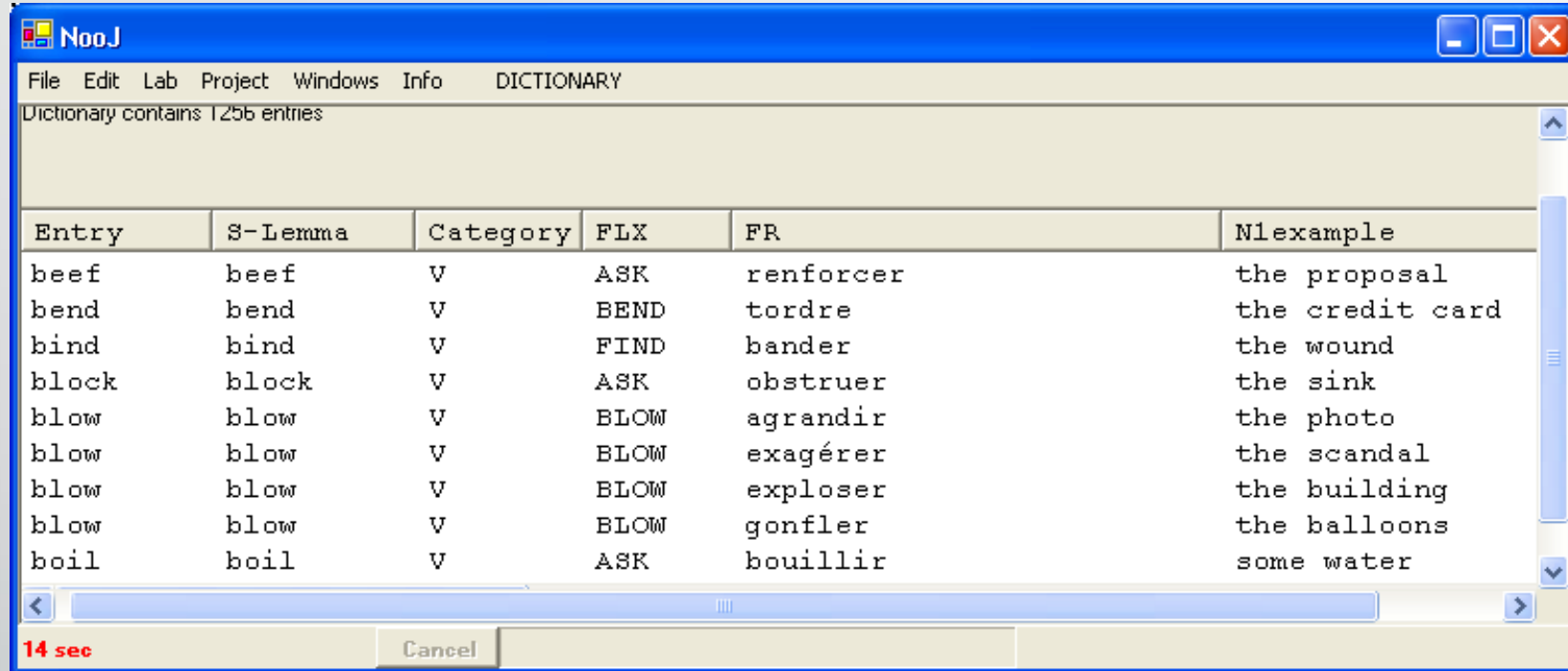
# English phrasal verbs (PV) or verb-particle combinations

- *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 13<sup>th</sup> to the 17<sup>th</sup> century, slight drop during in the 18<sup>th</sup> century, expansion in the 19<sup>th</sup> century
- often attributed to an American influence by grammarians
- “creativity 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

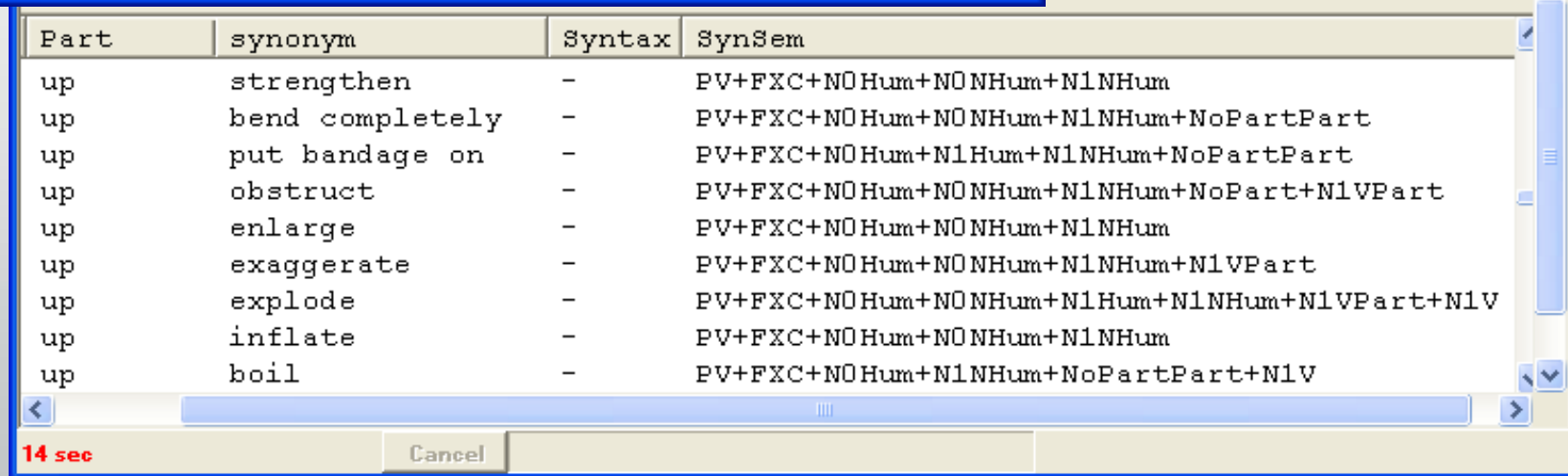
# NooJ PV Dictionary



The screenshot shows the NooJ application window with a menu bar (File, Edit, Lab, Project, Windows, Info, DICTIONARY) and a status bar indicating "Dictionary contains 1256 entries". The main content is a table with the following columns: Entry, S-Lemma, Category, FLX, FR, and N1example.

Entry	S-Lemma	Category	FLX	FR	N1example
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

At the bottom of the window, there is a red "14 sec" indicator and a "Cancel" button.

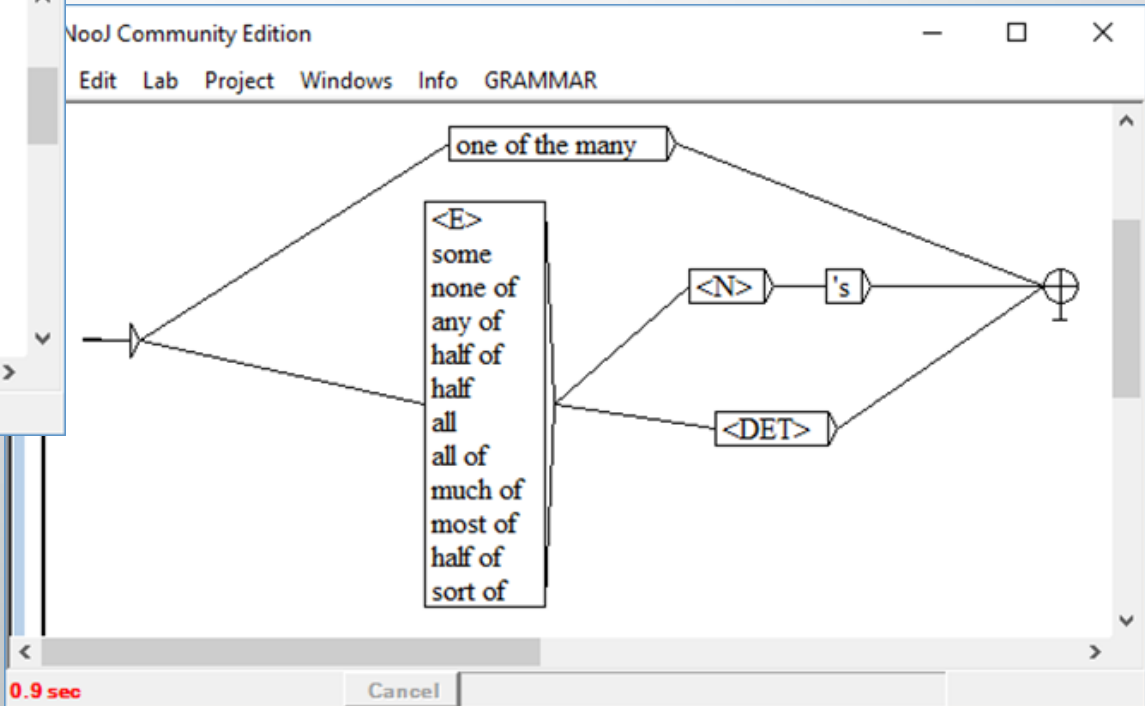
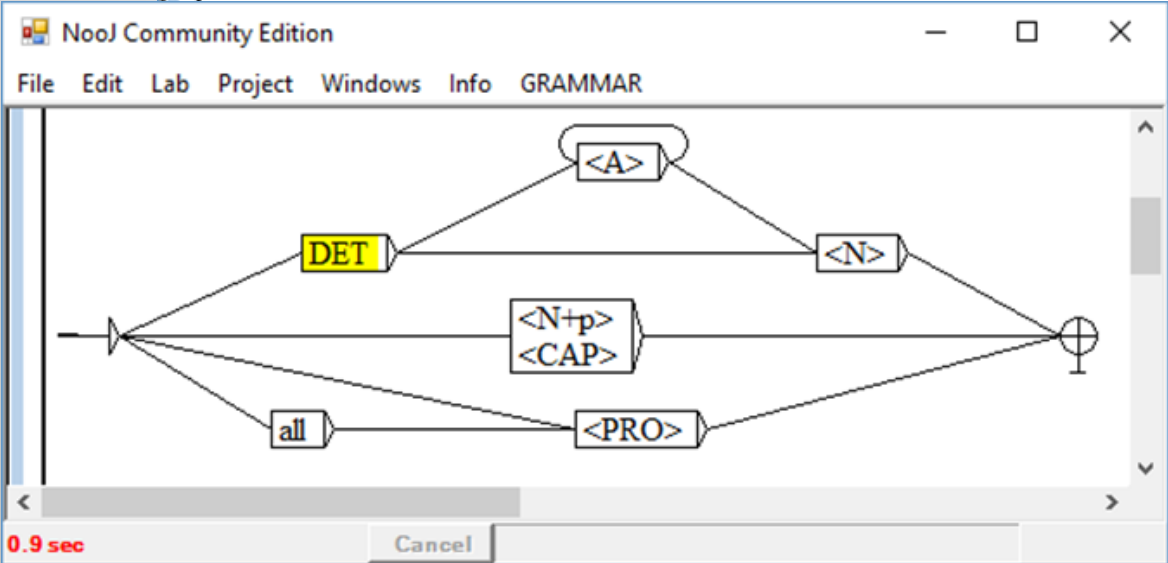
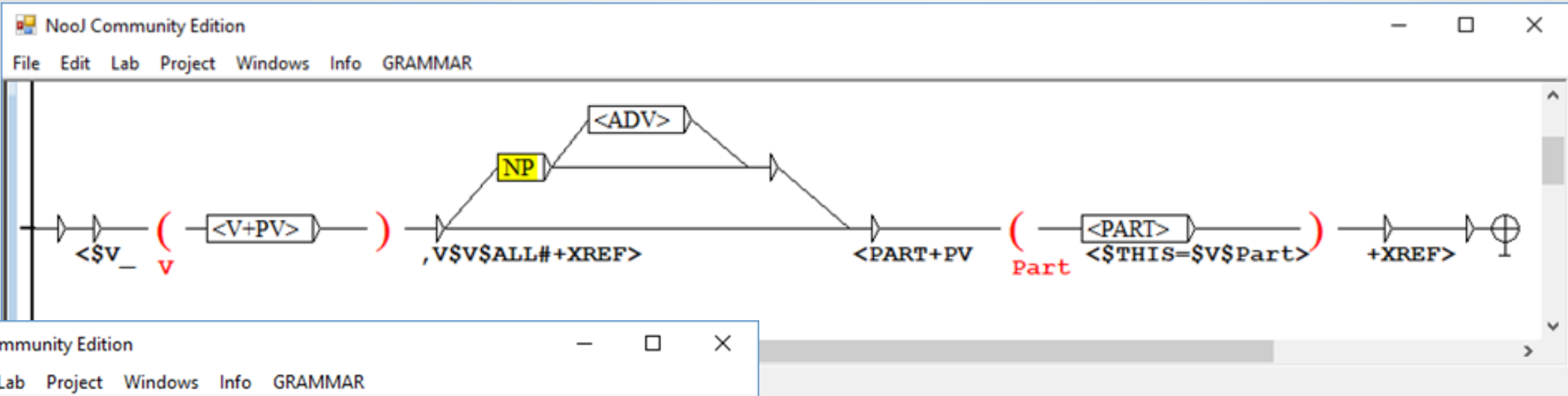


The screenshot shows a detailed view of a dictionary entry table with the following columns: Part, synonym, Syntax, and SynSem.

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

At the bottom of the window, there is a red "14 sec" indicator and a "Cancel" button.

# 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

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

⇒ 1881 novel *The Portrait of a Lady* (233,102 word forms)

⇒ \*PRECISION:  $\text{True Positives} / (\text{True Positives} + \text{False Positives})$



# 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 curse it! 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? In this spake Mr. Smangle, edging himself meanwhile as near as possible to the portmanteau, and beaming forth looks of the most fervent and disinterested friendship.

	204	208	211	217	224	231
=PP	any,DET	of,PREP	those,DET	little,A	thing,N+Nb=p	PART+PV
=PT+Pers=3+Nb=p	any,PRO		those,PRO	little,ADV+CmS=A		PART+PV
=PT+Pers=2+Nb=p						PART+PV
=PT+Pers=1+Nb=p						PART+PV
=PT+Pers=3+Nb=s						PART+PV
=PT+Pers=2+Nb=s						PART+PV
=PT+Pers=1+Nb=c						PART+PV

# 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)

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

\*PRECISION: True Positives/(True Positives + False Positives)

# RESULTS (Melville)

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

\*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