Using Electronic Dictionaries and NooJ to Generate Sentences Containing English Phrasal Verbs

Peter A. Machonis

Florida International University Department of Modern Languages, Miami, FL 33199, USA

machonis@fiu.edu

Abstract

This paper attempts to explore NooJ's "generation" mode to automatically produce transformations of sentences containing English Phrasal Verbs (PV). We exploit the same electronic dictionary and grammar previously used to recognize PV in large corpora (Machonis 2010, 2012), but have had to design a specific grammar for generating sentences, following the examples in Silberztein (2016), which showed how NooJ could generate over two million transformations or parallel sentences from the simple sentence Joe likes Lea. We created a grammar that can generate variations of a single phrase containing one of the PV found in the NooJ PV dictionary. For the moment the grammar only handles singular nouns in the present and past tense, but it is capable of applying a succession of transformations - particle movement, preterit, negation, clefting, modal insertion, aspect introduction, question formation, and passive voice, along with various combinations of these transformations - to over 1,200 PV from the electronic dictionary.

1 Introduction

English Phrasal Verbs (PV) have presented a fascinating challenge for Natural Language Processing, and as we will see in this paper, for Automatic Natural Language Generation, as well. We used the NooJ platform, a freeware linguistic development environment that can be downloaded from http://www.nooj4nlp.net/. NooJ allows linguists to describe several levels of linguistic phenomena and then apply formalized descriptions to any corpus of texts. Previously, we have used NooJ to identify all PV in large corpora, such as the complete novels of Dickens and Melville, other 19th novels, as well as a transcribed oral corpus of *Larry King Live* programs from January 2000.

Hodapp (2010), however, is the only researcher who has used NooJ to recognize PV and then apply the results for language generation. Using the NooJ PV grammar and dictionary, she designed a graphical user interface to help undergraduate students reduce PV usage – often considered informal – in academic papers. Her program generated single-word verb suggestions that could take the place of automatically identified PV.

This paper attempts to explore NooJ's "generation" mode to automatically produce paraphrases of sentences that are described by grammars. We exploit the same electronic dictionary used in NooJ to recognize PV, but have had to design a specific grammar for generating sentences that involve PV. As an initial experiment, we created a grammar that can generate variations of a single phrase containing a PV, involving transformations such as particle movement, preterit, negation, clefting - both of the subject and the object - modal insertion, aspect introduction, question formation, and passive voice, along with various combinations of these transformations. For example, from one simple sentence, such as Max figures out the problem, NooJ can generate over 2,500 variations such as Didn't Max figure out the problem?, It was Max who started to figure the problem out, He should figure it out, etc.

2 NooJ's PV Parsing Capabilities

Using NooJ, Machonis (2010, 2012) showed that the automatic recognition of PV proved to be far more complex than for other multi-word expressions due to three main factors: (1) their possible discontinuous nature (e.g., *let out the dogs \Leftrightarrow let the dogs out*), (2) their confusion with verbs followed by simple prepositions (e.g., *Do you remember what I asked you in Rome?* (preposition) vs. *Did you ask* the prince in when he arrived? (PV)), and (3) genuine ambiguity only resolvable from context (e.g., *Her neighbor was looking over the broken fence*, which can mean either "looking above the fence" (preposition) or "examining the fence" (PV)). On the bright side, though, NooJ can correctly identify many discontinuous PV, such as the following:

- (1) I **folded** all my bills **up** uniformly (*Great Expectations*)
- (2) he had that club-hammer there ... to **knock** some one's brains **out** with (*Moby Dick*)
- (3) a program that has effectively **brought** our crime rates **down** (*Larry King Live*).

NooJ requires both a grammar and a dictionary that work in tandem to annotate PV in large corpora. Figure 1 represents an example of NooJ's PV Grammar.



Figure 1: NooJ PV Grammar

The dictionary is based on previous work using Maurice Gross' (1994, 1996) Lexicon-Grammar approach. Lexicon-Grammar limits abstract notions in syntax and accentuates the reproducibility of linguistic data in the form of exhaustive syntactic tables, which are manually constructed and contain both lexical and syntactic information, as can be seen in the sample Table 1. From these Lexicon-Grammar tables of PV, we created a NooJ PV dictionary that contains more than 1,200 entries, which when used in tandem with the PV grammar, could automatically annotate PV in large corpora. Figure 2 is a sample of this dictionary, which mirrors much of the information contained within the Lexicon-Grammar entry seen in Table 1.

Although early experiments identified much noise, three disambiguation grammars, adverbial and adjectival expression filters, and idiom dictionaries were added to remove false PV without creating silence. This has made for a fairly intricate way to accurately annotate PV in large corpora.

Machonis (2016) explains how NooJ can successfully remove many inaccurate Text Annotation Structures (TAS).

$N_0 =: Nhum$	$N_0 =: N$ -hum	Verb	Particle	Example of N ₁	$N_1 =: Nhum$	$N_1 =: N\text{-}hum$	$N_0 V N_1$	N ₁ V Part	$N_1 V$	Synonym
+	+	beef	up	the proposal	-	+	-	-	-	strengthen
+	+	bend	up	the credit card	1	+	+	1	-	bend completely
+	I.	bind	up	the wound	+	+	+	I	-	bandage
+	+	block	up	the sink	1	+	+	+	-	obstruct
+	+	blow	up	the balloons	I.	+	I.	I.	-	inflate
+	+	blow	up	the building	+	+	I.	+	+	explode
+	+	blow	up	the photo	I.	+	I.	I.	-	enlarge
+	+	blow	up	the scandal	-	+	1	+	-	exaggerate
+	1	boil	up	some water	-	+	+	-	+	boil

Table 1: Sample from PV Lexicon-Grammar

Pie Ed. Lab. Poject Wedows Info DICTIONARY DEPENDENT LOSS OF THE SET OF	NooJ						
December Converter Like enter Batty S-Lemma Category FLX PR. Nlexample bard beef V ASK reforeer the proposal band bend V PTND bander the the crediter block block V ASK obstruer the sink block block V ASK obstruer the sink block blow V BLOW exgleser the sink blow blow V BLOW exgleser the ballens blow blow V BLOW exgleser the ballens blow blow V BLOW gonfler the ballens FRE SR Lab Popt Vndows bro DICHONARY FRE SR Lab Popt Vndows bro DICH	ie Edit Lab	Project Windows Info	o DICTIO	NARY			
Brtry S-Lemma Category FLX FR Niexample beef beef V ASK renforcer the proposal bend bend V BEND tordre the proposal bind bind V PIND bander the oregitsal block block V ASK obstruer the sink blow blow V BLOW agardir the photo blow blow V BLOW exagérer the scandal blow blow V BLOW exagérer the scandal blow blow V BLOW exagérer the balloons blow blow BLOW exagérer the balloons blow blow V BLOW exagérer the balloons blow blow V BLOW exagérer the balloons time blow V BLOW exagérer the balloons time Camel	ctionary contain	ns 1256 entries					
Brtry 8-Lemma Category FIX FR Niexample beef beef V ASK renforcer the proposal bend bend V BEND tordree the proposal bind bind V FIND bander the wound block block block block block the wound block block V ASK obstruer the wind blow blow V BLOW exgandir the photo blow blow V BLOW exgloser the bilding blow blow V BLOW gonder the bilding blow blow BLOW gonder the bilding far Concel PV+FXC+NDMm+NDMm+NIMMMM-NDFartPart pole							
basef beef V ASK renforcer the proposal bind bend V BEM todae the yroposal bind bind V PIND bander the wound block block V ASK obstruer the wound block block V ASK obstruer the wink blow blow V BLOW exadir the photo blow blow V BLOW gonfier the billons boil V BLOW gonfier the billons boil V BLOW gonfier the billons boil V BLOW gonfier the billons face Computer PV	Intry	S-Lemma	Category	FLX	FR	Nlexample	-
bend bend V BEND tordre the scelit ca: bind bind V PIND bander the sound block block block V ASK obstruer the sound block blow V BLOW exagérer the sound blow blow V BLOW exagérer the scandal blow blow V BLOW gondier the building blow blow V BLOW gonde gonde some water boil V ASK boullir some water some water some water frag East East Part Synonym DVFEXCHNOHum*NONHum*NNHum*NOPartPart some water up strongthen P PVFEXCHNOHum*NONHum*NNHum*NOPartPart some water up put bandage on PVFEXCHNOHum*NONHum*NNHum*NNPartPart some water some water up exagerate PVFEXCHNOHum*NONHum*NHum*NNMMAND*NPartPart some water <td>eef</td> <td>beef '</td> <td>v</td> <td>ASK</td> <td>renforcer</td> <td>the proposal</td> <td>_</td>	eef	beef '	v	ASK	renforcer	the proposal	_
bind bind V PIND bander the wound block block block obstruer the wind block block V ASK obstruer the sink blow blow V BLOW exgrandir the photo blow blow V BLOW exgloser the building blow blow V BLOW gonter the building blow V BLOW gonter the building some water file Edt boil V ASK bouillir some water file Edt Propt Wndws info <oltionary< td=""> some water file Edt Propt Wndws info<oltionary< td=""> some water exet Carreet PV#FXCHNNm+NDNTum+HINHum+NDFATEPart pV#FXCHNNm+NDNTum+HINHum+NDFATEPart up botruct - PV#FXCHNNm+NDNTum+HINHum+NDFATEPart pV#FXCHNNm+HINHum+HINHUMENNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN</oltionary<></oltionary<>	end	bend '	v	BEND	tordre	the credit card	
block block V ASK obstruer the sink blow blow V BLOW agrandir the photo blow blow V BLOW exagfer the photo blow blow V BLOW exagfer the photo blow blow V BLOW exagfer the balloons blow blow V BLOW exagfer the balloons boil V ASK bouilit some water some water test Cameed	ind	bind	v	FIND	bander	the wound	
blow blow V BLOW agrandir the photo blow blow V BLOW exgandir the photo blow blow V BLOW exgloser the building blow V BLOW gonfler the building blow V BLOW gonfler the building boil V ASK bouillir some water fm Edt boil V ASK bouillir some water fm Edt brogst tho OLTIONARY some water some water fm Edt bo Propt Wndows info OLTIONARY some water ext Syntax SynSem some water some water some water up strongthen P PV#F2CHNDHum+NDHTum+HINHum+NDFartPart up end completer PU#F2CHNDHum+NDHTum+HINHum+NDFartPart up optification PU#F2CHNDHum+HINHum+HINHum+NINFartHINFart PU#F2CHNDHum+HINHum+HINHUMPART some some som	lock	block '	v	ASK	obstruer	the sink	
<pre>slow blow v BLOW exgigrer the scandal low blow v BLOW gonfler the building slow blow v BLOW gonfler the building slow blow v BLOW gonfler the bullons soci boi v ASK bouillir some water the building some water some start the Edt Lob Projet Windows Info DKTIONARY and LOB Projet Windows Info DKTIONARY up strengthen - PV+FXC+NDRum+NDNHum+NIAHum PV+FXC+NDRum+NDNHum+NIAHum+NDPartPart up obstruct - PV+FXC+NDRum+NDNHum+NIAHum+NDPartPart up explode - PV+FXC+NDRum+NDNHum+NIAHum+NIAPart up explode - PV+FXC+NDRum+NDNHum+NIAHum+NIAPart up explode - PV+FXC+NDRum+NDNHum+NIAHum+NIAPart up explode - PV+FXC+NDRum+NDNHum+NIAHum+NIAPart up inflate - PV+FXC+NDRum+NDNHum+NIAHum winflate - PV+FXC+NDRum+NDNHum+NIAHum NAME Projet Part+NIAPART up boil - PV+FXC+NDRum+NDNHum+NIAPARTPart+NIAPART up boil - PV+FXC+NDRum+NDNHum+NIAPARTPart+NIAPARTPARTPART up boil - PV+FXC+NDRum+NDNHum+NIAPARTPARTPARTPARTPARTPARTPARTPARTPARTPAR</pre>	low	blow	v	BLOW	agrandir	the photo	
blow blow V BLOW exploser the building boil v ASK gonfler the building some water	low	blow	v	BLOW	exagérer	the scandal	
<pre>slow blow V BLOW gonfler the ballons oil boil V ASK bouillir some water ***********************************</pre>	low	blow	v	BLOW	exploser	the building	
boil V ASK bouillir some water 4 me	low	blow	v	BLOW	gonfler	the balloons	
Synopy Syntax SynBem Part synonym SynLax SynBem Part synonym SynLax SynBem Part synonym SynLax SynBem up strongthen - PV+FXC+ND1Rum+NDNHum+NLNHum up bed completely PV+FXC+ND1Rum+NDNHum+NLNHum+NoPartPart up betruct - PV+FXC+ND1Rum+NDNHum+NLNHum+NoPartPart up obstruct - PV+FXC+ND1Rum+NDNHum+NLNHum+NOPartPart up extgerate - PV+FXC+ND1Rum+NDNHum+NLNHum+NDPartPart up extgerate - PV+FXC+ND1Rum+NDNHum+NLNHum+NDPartPart up extgerate - PV+FXC+ND1Rum+ND1NHum+NLNHum+NDPartPart up extgerate - PV+FXC+ND1Rum+ND1NHum+NLNPart up extgerate - PV+FXC+ND1Rum+ND1NHum+NLNPart up extgerate - PV+FXC+ND1Rum+ND1NHum+NLNPart up extgerate - PV+FXC+ND1Rum+ND1NHum+NLNPart up inflate - PV+FXC+ND1Hum+ND1NHum+ND1NHum+NLNPart	oil	boil	v	ASK	bouillir	some water	
Same Control Mp Edit Lib Project Windows Info DCTIONARY ami Lib Project Windows Info DCTIONARY ami Lib Project Windows Info DCTIONARY ami Lib Project Windows Info DVFEXCHOINtam+NO INfram+NI NHam up strong them PVFEXCHOINtam+NO INfram+NI NHam up pub conclusion PVFEXCHOINtam+NO INfram+NI NHam+No PortPart up pub bandage on PVFEXCHOINtam+NO INfram+NI NHam up enlarge PVFEXCHOINtam+NO INfram+NI NHam <ni td="" vfart<=""> up inflate PVFEXCHOINtam+NO INfram+NI NHam<ni td="" vfart+ni="" vfart<=""> up inflate PVFEXCHOINtam+NI NHAm<ni nham<="" td=""></ni></ni></ni>							5
Synopy Syntax Syn3em Part synonym Syn3em Part synonym Syn3em up strongthen - PV+FXC+NOHum+NONHum+NINHum PVeFXC+NOHum+NONHum+NINHum+NoPartPart up bstruct - PV+FXC+NOHum+NONHum+NINHum+NoPartPart - up bstruct - PV+FXC+NOHum+NONHum+NINHum+NoPartPart - up bstruct - PV+FXC+NOHum+NONHum+NINHum+NoPartPart - up estruct - PV+FXC+NOHum+NONHum+NINHum+NUPartPart - up estruct - up estruct - PV+FXC+NOHum+NOHHum+NUHM=NVPartPart - up estruct - up <td></td> <td></td> <td>. [</td> <td></td> <td></td> <td></td> <td>-</td>			. [-
By Construct Synthesis Syntax SynSem Part synonym SynSem Part synonym SynSem up strongthen - PV+FXC+NOHum+NONHum+NINHum up bend completely PV+FXC+NOHum+NONHum+NINHum+NoPartPart up put bandage on - PV+FXC+NOHum+NONHum+NINHum+NoPartPart up obstruct - PV+FXC+NOHum+NONHum+NONHum+NoPartPart up extgerate - PV+FXC+NOHum+NONHum+NOHUM=NOPartPart up extgerate - PV+FXC+NOHum+NOHHum+NINHumN=NOPartPart up extgerate - PV+FXC+NOHum+NOHHum+NINHumN=NOPartPart up extgerate - PV+FXC+NOHum+NOHHum+NINHum+NVPart up extgerate - PV+FXC+NOHum+NOHHum+NINHum+NVPart up extgerate - PV+FXC+NOHum+NOHNHum+NINHum+NVPart up extgerate - PV+FXC+NOHum+NOHNHum+NINHum+NVPart+NUPart up extgerate - PV+FXC+NOHum+NOHNHum+NINHum+NIVPart+NUPart+NI up inflate - PV+FXC+NO							
He Ed. tob Project Windows Info DICTIONARY Gara T.258 errors Part Synonym Byntax SynSem up strongthon - PV+FXC+ND1kum+ND1NHum+NLMHum up bend completely > PV+FXC+ND1kum+ND1NHum+NLMHum+NoPartPart up put bandage on - PV+FXC+ND1kum+ND1NHum+NLMHum+NoPartPart up extragerate > PV+FXC+ND1kum+ND1NHum+NLMHum-NoPartPart up extragerate - PV+FXC+ND1kum+ND1NHum+NLMHum-NoPartPart up extragerate > PV+FXC+ND1kum+ND1NHum+NLMHum-ND2rt+NIZ up extragerate - PV+FXC+ND1kum+ND1NHum+NLMHum+NLVPart up extragerate - PV+FXC+ND1kum+ND1NHum+NLNHum+NLVPart up extragerate - PV+FXC+ND1kum+ND1NHum+NLVPart up inflate - PV+FXC+ND1kum+ND1NHum+NLNHum+NLVPart+NLVPart up inflate - PV+FXC+ND1kum+ND1NHum+NLNHum	NooJ						
Part Synonym SynSem up stengthen PV+FXC+NDH.m+HONKum+HIMum up bend completely PV+FXC+NDH.m+HONKum+HIMum+NoPartPart up bend completely PV+FXC+NDH.mm+HONKum+HIMMam+NoPartPart up obstruct PV+FXC+NDH.mm+HONKum+HIMMam+NoPartPart up obstruct PV+FXC+NDH.mm+HONKum+HIMMam*NoPartPart up exaggerate PV+FXC+NDH.mm+HONKum+HIMMam*NoPartPart up exaggerate PV+FXC+NDH.mm+HONKum+HIMMam*NVPart up inflate PV+FXC+NDH.mm+HONKum+HIMMam*NVPart+NVPart+NV up boil PV+FXC+NDH.mm+HONKum+HIMMam	le Edit Lab	Project Windows In	fo DICTI	ONARY			
Part synonym SynEax SynEax up strongthon - PV+FXC+N0Hum+N0NHum+N1NHum up bond completely - PV+FXC+N0Hum+N0NHum+N1NHum+NoPartPart up bottout - PV+FXC+N0Hum+N0NHum+N1NHum+NoPartPart up obstruct - PV+FXC+N0Hum+N0NHum+N1NHum+NoPartPart up exagerate - PV+FXC+N0Hum+N0NHum+N1NHum+NoPartPart up exagerate - PV+FXC+N0Hum+N0NHum+N1NHum+NoPartPart up explorde - PV+FXC+N0Hum+N0NHum+N1NHum+N1VPart up explorde - PV+FXC+N0Hum+N0NHum+N1NHum+N1VPart up explorde - PV+FXC+N0Hum+N0NHum+N1NHum+N1VPart up inflate - PV+FXC+N0Hum+N0NHum+N1NHum+N1VPart+N1VPart up inflate - PV+FXC+N0Hum+N0NHum+N1NHum+N1VPart+N1VPart up inflate - PV+FXC+N0Hum+N0NHUM+N1NHum	ins 1256 entrie	95					
Part synonym Syntax SynSem up strengthen - PV+F2C+ND1Hum+HD1HHum+M1HHum up bend completed - PV+F2C+HD1Hum+HD1HHum+M1HHum+NoPartPart up put bandage on - PV+F2C+HD1Hum+HD1HHum+HD1HHum+NoPartPart up obstruct - PV+F2C+HD1Hum+HD1HHum+N1HHum+NoPartPart up alarge - PV+F2C+HD1Hum+HD1HHum+N1HHum+NoPart+N1VPart up exaggerste - PV+F2C+HD1Hum+HD1HHum+N1HHum+N1VPart up exajded - PV+F2C+HD1Hum+HD1HHum+HD1HHum+N1VPart up inflate - PV+F2C+HD1Hum+HD1HHum+HD1HHum+HD1HHum+HD1VPart up inflate - PV+F2C+HD1Hum+HD1HHum+HD1HHum							
Part synony Syntax Syndam up strongthon PV+FXC+N0Hnm+N0NHum+X1HHum up bend completely - PV+FXC+N0Hnm+N0NHum+X1HHum+NoPartPart up put bandage on - PV+FXC+N0Hnm+N0NHum+X1NHum+NoPartPart up obstruct - PV+FXC+N0Hnm+N0NHum+X1NHum+NoPartPart up enlarge - PV+FXC+N0Hnm+N0NHum+X1NHum+NoPartPart up exaggerate - PV+FXC+N0Hnm+N0NHum+X1NHum+NoPartPart up explode - PV+FXC+N0Hnm+N0NHum+X1NHum+NoPartPart+N1V up explode - PV+FXC+N0Hnm+N0NHum+N1NHum+N1VPart+N1V up explode - PV+FXC+N0Hnm+N0NHum+N1VPart+N1V up explode - PV+FXC+N0Hnm+N0NHum+N1VPart+N1V up explode - PV+FXC+N0Hnm+N0NHum+N1VPart+N1V up boil - PV+FXC+N0Hnm+N0NHum+N0NF							
up strongthen PV+2X <h01hum+m1nhum+n1nhum< th=""> up bend completed PV+2X<h01hum+m1nhum+n1nhum+n0partpart< td=""> up put bandage on PV+2X<h01hum+n1nhum+n1nhum+n0partpart< td=""> up betruct PV+2X<h01hum+n1nhum+n1nhum+n0partpart< td=""> up enlarge PV+2X<h01hum+n1nhum+n1nhum+n1partpart< td=""> up enlarge PV+2X<h01hum+n1nhum+n1nhum+n1part< td=""> up explode PV+2X<h01hum+n1nhum+n1nhum+n1vpart+n1< td=""> up explode PV+2X<h01hum+n1nhum+n1nhum+n1vpart+n1< td=""> up inflate PV+2X<h01hum+n1nhum+n1nhum+n1vpart+n1< td=""> up boll PV+2X<h01hum+n1nhum+n1nhum+n1vpart+n1< td=""></h01hum+n1nhum+n1nhum+n1vpart+n1<></h01hum+n1nhum+n1nhum+n1vpart+n1<></h01hum+n1nhum+n1nhum+n1vpart+n1<></h01hum+n1nhum+n1nhum+n1vpart+n1<></h01hum+n1nhum+n1nhum+n1part<></h01hum+n1nhum+n1nhum+n1partpart<></h01hum+n1nhum+n1nhum+n0partpart<></h01hum+n1nhum+n1nhum+n0partpart<></h01hum+m1nhum+n1nhum+n0partpart<></h01hum+m1nhum+n1nhum<>	Part	synonym		Syntax	SynSem		1
up bend completely PV+FXC+N0Hum+N0Hum+N1Hum+N1Mem*N0PartPart up obstruct PV+FXC+N0Hum+N0Hum+N1Hum+N1Hum+N1Part up enlarge PV+FXC+N0Hum+N0Hum+N1Hum+N1Hum+N1Part up enlarge PV+FXC+N0Hum+N0Hum+N1Hum+N1Hum+N1Part up explode PV+FXC+N0Hum+N0Hum+N1Hum+N1Hum+N1Part up explode PV+FXC+N0Hum+N1Hum+N1Hum+N1VPart up explode PV+FXC+N0Hum+N1Hum+N1Hum+N1VPart+N1 up explode PV+FXC+N0Hum+N1Hum+N1Hum+N1VPart+N1VPart+N1 up boil PV+FXC+N0Hum+N1Hum+N1Hum+N1VPart+N1	up	strengthen		-	PV+FXC+N0Hum+N0NHur	+N1NHum	
up put bandage on PV+FXC+NDHum+NINHum+NDFactPact up obstruct - PV+FXC+NDHum+NDHTum+NINHum+NDFactPact up enlarge - PV+FXC+NDHum+NDHTum+NINHum+NDFactPact up exggerate - PV+FXC+NDHum+NDHTum+NINHum+NINHum+NIVFact up explode - PV+FXC+NDHum+NDHTum+NINHum+NIVFact+NI up inflate - PV+FXC+NDHum+NDHTum+NINHum+NIVFact+NI up boll - PV+FXC+NDHum+NDHTum+NINHUF	up	bend comple	tely ·	-	PV+FXC+N0Hum+N0NHur	n+N1NHum+NoPartPart	
up obstruct PV+FXC+NDHum+NDHum+NIHHum+NoPart+NIVPart up enlarge - PV+FXC+NDHum+NOHHum+NIHHum up exsgerate - PV+FXC+NDHum+NOHHum+NIHHum up exsplore - PV+FXC+NDHum+NOHHum+NIHHum up explore - PV+FXC+NDHum+NOHHum+NIHHum+NIVPart up inflate - PV+FXC+NDHum+NOHHum+NIHHum+NIVPart+NI up boil - PV+FXC+NDHum+NOHHum+NIHHum+NIVPart+NIV	up	put bandage	on	-	PV+FXC+N0Hum+N1Hum-	N1NHum+NoPartPart	
up enlarge - PV+FXC+NDHum+NLINHum up exsgerate - PV+FXC+NDHum+NDINHum+NLINHum NLIVPart up explode - PV+FXC+NDHum+NDINHum+NLINHum NLIVPart+NL up inflate - PV+FXC+NDHum+NDINHum+NLINHum+NLIVPart+NL up inflate - PV+FXC+NDHum+NDINHum+NLNHUm up boll - PV+FXC+NDHum+NDINHum+NDFart+PLart+NLV	up	obstruct		-	PV+FXC+N0Hum+N0NHur	1+N1NHum+NoPart+N1VPart	
up exaggerate PV+FXC+NDHum+NDHum+NIVPart up explode - PV+FXC+NDHum+NDHum+NLVPart+NL up inflate - PV+FXC+NDHum+NDHUm+NLHMm up boil - PV+FXC+NDHum+NDHUm+NLHMm	up	enlarge		-	PV+FXC+N0Hum+N0NHur	1+N1NHum	
up explode - PV+FXC+NDHum+NDHVum+NLHum+NLVPart+NL up inflate - PV+FXC+NDHum+NDHVum+NLHum up boil - PV+FXC+NDHum+NDHVum+NLHum	up	exaggerate		-	PV+FXC+N0Hum+N0NHur	1+N1NHum+N1VPart	
up inflate - PV+FXC+NOHum+NONHum+N1NHum up boil - PV+FXC+NOHum+N1NHum+NoPartPart+N1V	up	explode		-	PV+FXC+N0Hum+N0NHur	+N1Hum+N1NHum+N1VPart+N1V	
up boil - PV+FXC+N0Hum+N1NHum+NoPartPart+N1V		0.00 0.00 0.00 0					
	up	inflate		-	PV+FXC+N0Hum+N0NHur	x+N1NHum	
	up	inflate boil		-	PV+FXC+N0Hum+N0NHur PV+FXC+N0Hum+N1NHur	1+N1NHum 1+NoPartPart+N1V	
A rea	up up	inflate boil		-	PV+FXC+NOHum+NONHur PV+FXC+NOHum+N1NHur	u+NiNHum u+NoPartPart+N1V	~ ~

Figure 2: NooJ PV Dictionary

Overall, our NooJ PV studies have achieved 88% accuracy, with most of the noise coming from the particles in and on, which are fairly tricky to distinguish automatically from prepositions (e.g. had a strange smile on her thin lips (preposition) vs. had her hat and jacket on (PV)). However, in a more recent study on the novels of Dickens and Melville, we reduced the NooJ dictionary to include only six particles (out, up, down, away, back, off) instead of twelve, which helped us achieve 98% accuracy. Other linguists, such as Hiltunen (1994:135), also limited searches to these six typical particles representing three levels of PV frequency: high (out, up), mid (down, away), and low (back, off). However, for our generation study, we used the original PV dictionary of over 1,200 entries.



Figure 3: NooJ PV Generate Grammar

3 NooJ's PV Generating Capabilities

As an initial experiment, we created a grammar that can generate variations of a single phrase containing a PV. We changed the local variables in the original PV parsing grammar to global variables, and then added new pathways for the various transformations, following the examples in Silberztein (2015, 2016), which showed how NooJ could generate over two million transformations or parallel sentences from simple sentences, such as *Luc aime Léa* 'Luke likes Lea' and *Joe likes Lea*.

For the moment the grammar (Figure 3) only handles singular nouns in the present and past tense, but it is capable of applying a succession of transformations – particle movement, preterit, negation, clefting (both of the subject and the object), modal insertion, aspect introduction, question formation, and passive voice – to over 1,200 PV from the electronic dictionary. The upper pathway recognizes a sentence containing a PV and creates the first variants, with past tense and particle movement. So, if we enter: *Max figures out the problem*, that sentence is recognized, but NooJ also creates, *Max figured out the problem, Max figures the problem out*, and *Max figured the problem out*. Also, *Max* and *the problem* could be substituted by pronouns.

The next series of transformations deal with negation, modal insertion and aspect introduction. These pathways create variations such as *Max did* not figure out the problem, Max could figure out the problem, Max started to figure the problem out, Max finished figuring out the problem, etc. Our modal subgraph includes nine modal verbs – can, could, may, might, must, should, ought to, need to, have to -- as well as negative variants and contracted forms, such as can't, shouldn't, etc. Our aspectual variants include inchoative (begin, start), durative (continue), and completive (finish, stop), with both negative and preterit possibilities.

In the middle of the graph, we have the clefting conduit, either *It is he who*, *It is Max who*, or *It is the problem that*, followed by the same PV. In the case of clefting of the subject, the sentence will also undergo particle movement, negation, modal insertion, and aspectual variants. Thus sentences such as the following would be created: *It was Max who didn't figure it out*, *It is Max who may not figure the problem out*, *It was Max who started to figure the problem out*, *It was Max who didn't finish figuring it out*, etc.

The final two pathways involve question formation and passive voice formation: *Does Max figure out the problem*? and *The problem is figured out by Max*. These variants are also open to negative and preterit transformations, along with pronominal forms, such as: *Didn't Max figure it out*?, *The problem was not figured out by him*, etc.

Bob figured the problem out	SENT+Preterit+PartMvt				
Bob has to figure it out	SENT+ModHave+Pro1				
It is Bob who could figure out the problem	SENT+Cleft0+ModCan+Preterit				
Mike must clear away the area	SENT+ModMust				
Mike needn't clear the area away	SENT+ModNeed+Neg+Contraction+PartMvt				
Mike oughtn't clear it away	SENT+ModOught+Neg+Contraction+Pro1				
Brent continued to push the deadline back	SENT+Preterit+AspDurative+PartMvt				
Didn't Brent push the deadline back ?	SENT+Question+Preterit+Neg+Contraction+PartMvt				
It is not he who pushed the deadline back	SENT+Neg+Pro0+Cleft0+Preterit+PartMvt				
Did Blake hand the exam in ?	SENT+Question+Preterit+PartMvt				
He handed it in	SENT+Pro0+Preterit+Pro1				
It wasn't Blake who handed in the exam	SENT+Preterit+Neg+Contraction+Cleft0+Preterit				
Phil might call off the trip	SENT+ModMay+Preterit				
Phil started to call the trip off	SENT+Preterit+AspInchoative+PartMvt				
It isn't Phil who called it off	SENT+Neg+Contraction+Cleft0+Preterit+Pro1				
Steve started turning on the radio,	SENT+Preterit+AspInchoative				
Steve began to turn the radio on,	SENT+Preterit+AspInchoative+PartMvt				
It wasn't Steve who didn't turn the radio on	SENT+Preterit+Neg+Contraction+Cleft0+Preterit+Neg+Contraction+PartMvt				
Max couldn't burn the building down	SENT+ModCan+Preterit+Neg+Contraction+PartMvt				
Max stops burning down the building	SENT+AspCompletive				
The building wasn't burnt down by Max	SENT+Passive+Preterit+Neg+Contraction				
Didn't Devon shred the document up ?	SENT+Question+Preterit+Neg+Contraction+PartMvt				
Did Devon shred it up ?	SENT+Question+Preterit+Pro1				
It isn't Devon who couldn't shred it up	SENT+Neg+Contraction+Cleft0+ModCan+Preterit+Neg+Contraction+Pro1				

Table 2: Sample Sentences from NooJ PV Generate Grammar with Transformations Noted

All in all, for every sentence containing a PV, this NooJ grammar will create 2,694 entries, and sometimes more in the case of certain irregular verbs, such as *burn*, which has two past tenses, *burned* and *burnt*. This grammar can also generate sentences based on all the other PV in the NooJ PV dictionary such as, *clear away the area*, *push back the deadline*, *hand in the exam*, *call off the trip*, *turn on the radio*, *burn down the building*, *shred up the document*, etc. as can be seen in Table 2. If we apply the PV Generate Grammar to all of the 1,200 verbs listed in the NooJ PV dictionary, NooJ would have generated over three million different sentences.

4 Conclusion

Not only does this research shed light on a major NLG problem, i.e., generating sentences containing discontinuous multiword expressions, but it seems to approach solving the original Chomskian challenge of generating "all and only" the sentences of a language. Some of these sentences might sound more natural than others, and some will need a specific context to appear likely, however, all of the sentences generated are grammatical. Speakers may choose certain forms over others during the course of a conversation, in what might be called discourse management. Nevertheless, NooJ does allow the user to specify which transformations are to be applied and thus limit the number of sentences generated to a specific context. As can be seen in this preliminary test, NooJ is a very powerful tool for linguistics, as well as Natural Language Generation.

References

- Maurice Gross. 1994. Constructing Lexicon-Grammars. *Computational Approaches to the Lexicon* (eds. Atkins and Zampolli), 213-263. Oxford University Press, Oxford, UK
- Maurice Gross. 1996. Lexicon Grammar. Concise Encyclopedia of Syntactic Theories (eds. K. Brown and J. Miller), 244-258. Elsevier, New York.
- Risto Hiltunen. 1994. On Phrasal Verbs in Early Modern English: Notes on Lexis and Style. *Studies in Early Modern English* (ed. D. Kastovsky), 129-140. Mouton de Gruyter, Berlin, Germany.
- Lien Huynh Hodapp. 2010. *Eliminating phrasal verbs in academic writing*. Minnesota State University Memorial Library University Archives, Mankato, Minnesota.
- Peter A. Machonis. 2010. English Phrasal Verbs: from Lexicon-Grammar to Natural Language Processing. *Southern Journal of Linguistics* 34(1): 21-48.

- Peter A. Machonis. 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 (eds K. Vučković, B. Bekavac, and M. Silberztein), 152-165. Cambridge Scholars Publishing, Newcastle upon Tyne, UK.
- Peter A. Machonis. 2016. Phrasal Verb Disambiguating Grammars: Cutting Out Noise Automatically. Automatic Processing of Natural-Language Electronic Texts with NooJ. Communications in Computer and Information Science book series (CCIS, volume 667), (eds L. Barone, M. Monteleone and M. Silberztein). 169-181. Springer International Publishing AG, Cham, Switzerland.
- Max Silberztein. 2015. La formalisation des langues: l'approache de NooJ. ISTE Editions, London,UK.
- Max Silberztein. 2016. Formalizing Natural Languages: The NooJ Approach. Wiley ISTE, London, UK.