

Who is afraid of Phrasal Verbs:

the case of a learner corpus

Joanna Wierszycka

Joanna Wierszycka

Who is afraid of Phrasal Verbs:
the case of a learner corpus

© Copyright by Joanna Wierszycka & e-bookowo
Projekt finansowany ze środków Narodowego Centrum Nauki
z grantu MNiSW nr N N104 378740
Grafika na okładce: e-bookowo
Projekt okładki: e-bookowo

ISBN 978-83-7859-502-1

Wydawca: Wydawnictwo internetowe e-bookowo
www.e-bookowo.pl
Kontakt: wydawnictwo@e-bookowo.pl

Wszelkie prawa zastrzeżone.
Kopiowanie, rozpowszechnianie części lub całości
bez zgody wydawcy zabronione
Wydanie I 2015

'It's a special form of scholarly neurosis ... no longer able to distinguish between life and literature.'
(David Lodge, *The British Museum Is Falling Down*, 63)

To Ula

Table of contents

TABLE OF CONTENTS.....	5
LIST OF TABLES.....	9
LIST OF FIGURES.....	11
LIST OF ABBREVIATIONS.....	12
INTRODUCTION.....	15
CHAPTER 1: THEORETICAL APPROACHES TO PHRASAL VERBS AND LEARNER STUDIES.....	18
1.1. INTRODUCTION.....	18
1.2. THE STRUCTURE OF PHRASAL VERBS.....	19
1.3. THEORETICAL VERIFICATION OF THE PHRASAL VERB STATUS.....	20
1.3.1. <i>Bolinger (1971)</i>	22
1.3.2. <i>Darwin and Gray (1999)</i>	26
1.3.3. <i>Quirk et al. (1985)</i>	31
1.3.4. <i>Biber et al. (1999)</i>	32
1.3.5. <i>Rudzka-Ostyn (2003)</i>	35
1.3.6. <i>Armstrong (2004)</i>	37
1.4. SELECTED APPROACH: DARWIN AND GRAY (1999).....	39
1.5. SEMANTICS.....	40
1.5.1. <i>The particle</i>	40
1.5.2. <i>The metaphor behind the particle</i>	43
1.5.3. <i>Semantic categories of phrasal verbs: compositionality</i>	46
1.6. PHRASAL VERBS AS A PHENOMENON FROM THE REALM OF SPEECH.....	49

1.7. PHRASAL VERBS AND LEARNERS.....	51
1.7.1. <i>Avoidance</i>	51
1.7.2. <i>Other studies on phrasal verbs and learners</i>	53
1.7.3. <i>Phrasal verbs and the Polish Learner</i>	54
1.7.4. <i>Teaching phrasal verbs</i>	56
CHAPTER 2: LANGUAGE CORPORA AND LINGUISTIC DATA ANNOTATION.....	58
2.1. INTRODUCTION: THE CORPUS PHILOSOPHY.....	58
2.1.1. <i>Corpus-based vs. corpus-driven</i>	60
2.2. TYPES OF CORPORA.....	61
2.2.1. <i>Spoken corpora</i>	64
2.2.2. <i>Learner corpora</i>	67
2.2.3. <i>Spoken learner corpora</i>	70
2.3. DATA ANNOTATION.....	70
2.3.1. <i>Types of linguistic annotation</i>	71
2.3.2. <i>Learner corpus annotation</i>	78
2.4. CONCLUDING REMARKS.....	80
CHAPTER 3: POS-TAGGING OF POLISH LINDSEI CORPUS.....	82
3.1. INTRODUCTION.....	82
3.2. THE LINDSEI CORPUS.....	83
3.2.1. <i>The defining features of the LINDSEI corpus</i>	84
3.2.2. <i>The LINDSEI design criteria</i>	85
3.2.3. <i>Learner variables: social</i>	87
3.2.4. <i>Learner variables: FLL-related</i>	88
3.2.5. <i>Task variables</i>	88
3.2.6. <i>Topic distribution</i>	89
3.2.7. <i>Transcriptions</i>	90
3.2.8. <i>Hands-on transcribing</i>	90
3.3. POS-TAGGING: THE PROPER TAGGER.....	93
3.3.1. <i>The original CLAWS tagger and its operating principles</i>	95
3.4. PROBLEMS WITH TAGGING PLINDSEI.....	98
3.4.1. <i>The spelling problem</i>	99

3.4.2. <i>Transcribing conventions</i>	100
3.4.3. <i>Speech-related issues</i>	102
3.4.4. <i>Learner-related issues</i>	103
3.4.5. <i>Correctly POS-tagged learner errors</i>	104
3.4.6. <i>Tagger inconsistency</i>	105
3.5. INCREASING TAGGER ACCURACY: PROBLEMATIC AREAS IDENTIFIED.....	107
3.5.1. <i>Spoken CLAWS version</i>	109
3.5.2. <i>Transcription format</i>	109
3.6. PRACTICAL SOLUTIONS: THE IDIOMLIST MODULE AT WORK.....	110
3.7. PRACTICAL SOLUTIONS: LEXICAL POS-TAGGER RESOURCES.....	114
3.8. RESULTS AND CONCLUSIONS.....	116

CHAPTER 4: PHRASAL VERBS IN LEARNER ENGLISH: GENERAL AND SEMANTIC CLASSIFICATION.....119

4.1. INTRODUCTION.....	119
4.2. RESEARCH QUESTIONS.....	120
4.3. CL METHODOLOGY: CA, CIA AND VARIABLES.....	122
4.3.1. <i>PLINDSEI and LOCNEC: comparability and statistical significance</i>	127
4.4. TOOLS.....	128
4.5. DATA EXTRACTION PROCEDURES.....	129
4.5.1. <i>WordSmith Tools</i>	130
4.5.2. <i>Holistic searches</i>	133
4.6. EMPIRICAL VERIFICATION OF THE PHRASAL VERB STATUS.....	135
4.6.1. <i>First filter</i>	136
4.6.2. <i>Second filter</i>	137
4.6.3. <i>Outcome of context-tag searches on NS data</i>	138
4.6.4. <i>Outcome of context-tag searches on PL data</i>	139
4.6.5. <i>Counting phrasal verbs</i>	141
4.7. RESEARCH QUESTION NUMBER ONE: THE QUANTITATIVE RESULTS. DISCUSSION OF FINDINGS	142
4.8. RESEARCH QUESTION NUMBER TWO: THE SEMANTIC APPROACH TO PV GROUPING.....	143
4.8.1. <i>Stage one (the particle)</i>	143
4.8.2. <i>Stage two (compositionality)</i>	145

4.8.3. <i>Proportion of idiomatic (to non-idiomatic) PVs</i>	150
4.8.4. <i>Compositionality: a rising trend?</i>	151
4.8.5. <i>The influence of setting and interviewer on the data recorded</i>	155
4.9. RESEARCH QUESTION NUMBER THREE: INDIVIDUAL LEARNER TENDENCIES.....	156
4.10. ASSUMPTIONS ABOUT THE LOW LEARNER PV USE.....	160
4.10.1. <i>Phrasal verbs in coursebooks</i>	161
4.10.2. <i>Phrasal verbs: Finding better equivalents</i>	162
4.11. VERIFICATION OF RESEARCH QUESTIONS AND CONCLUDING REMARKS.....	164
CONCLUSIONS	167
REFERENCES	170
LIST OF APPENDICES	188

List of tables

Table 1: 9 tests for Phrasal Verb verification and problems therein (after Bolinger 1971: 6-17).....	22
Table 2: Darwin and Gray's (1999) 7 tests for Phrasal Verb verification.....	28
Table 3: Principal types of multi-word verbs (after Quirk et al. 1985: 1161).....	32
Table 4: Multi-word lexical verbs (after Biber et al. 1999: 403-405).....	34
Table 5: Compositionality categories in Phrasal Verbs (Armstrong 2004: 222).....	38
Table 6: Particle meanings (after Rudzka-Ostyn 2003: 14-204).....	43
Table 7: Sets of metaphors (after Moon 2005).....	45
Table 8: PV compositionality.....	46
Table 9: A sentence tagged with the TOSCA POS tagger.....	71
Table 10: LINDSEI partners/contributors (adapted from Gilquin et al. 2010).....	85
Table 11: Number of words in Polish, French and native English LINDSEI subcorpora interviews.....	87
Table 12: Major error-prone areas after initial run with CLAWS on PLINDSEI.....	108
Table 13: An extract from Appendix G.....	116
Table 14: Research questions concerning learner PV use.....	122
Table 15: CEF rating of LINDSEI sample (Gilquin et al. 2010: 11).....	126
Table 16: General word and PV token counts in PL and NS corpora.....	128
Table 17: C7 tags used for searching PVs.....	131
Table 18: Phrasal Verb general token counts.....	142
Table 19: Phrasal Verb general type counts + TTR.....	142
Table 20: Richness of particles in NS and PL speakers.....	144
Table 21: Polysemy across speaker groups: Common PVs with different senses.....	144

Table 22: Selected Phrasal Verb forms found across compositionality groups in the PLINDSEI corpus.....	147
Table 23: Counting Phrasal Verbs: the example of the particle around from the PLINDSEI corpus.....	148
Table 24: Selected Phrasal Verb forms found across compositionality groups in the LOCNEC corpus.....	149
Table 25: Categories of PV compositionality in PL and NS corpora, expressed in PV types.....	150
Table 26: Categories of PV compositionality in PL and NS corpora, expressed in PV tokens.....	150
Table 27: Categories of PV compositionality in PL and NS corpora, expressed in phrasal verb TTR.....	150
Table 28: Categories of PV compositionality in PL and NS corpora, expressed in PV types.....	152
Table 29: Chi-square calculations for PV compositionality types in PL and NS corpora	152
Table 30: Categories of PV compositionality in PL and NS corpora, expressed in PV tokens.....	153
Table 31: Chi-square calculations for PV compositionality tokens in PL and NS corpora	153
Table 32: Categories of PV compositionality in PL and NS corpora, expressed in phrasal verb TTR.....	153
Table 33: Number of PVs in relation to length of stay in UK and number of years of English at school.....	157
Table 34: PV tokens: Individual user perspectives compared.....	158
Table 35: Research questions concerning learner PV answered.....	166

List of figures

Figure 1: Defining the Phrasal Verb (after Darwin and Gray 1999: 68).....	26
Figure 2: Metaphorical or extended meanings of the particles (Rudzka-Ostyn 2003: 4).....	42
Figure 3: The trend of idiomatic compositionality.....	49
Figure 4: A screenshot of the SoundScriber freeware program.....	91
Figure 5: Contrastive interlanguage analysis (after Granger 1996: 44).....	123
Figure 6: ICLE general and L2-specific variables (after Granger 2004: 126).....	124
Figure 7: LINDSEI variables (after Gilquin et al. 2010: 7).....	125
Figure 8: WordSmith Tools (WST) main window.....	130
Figure 9: Concord: Advanced settings.....	131
Figure 10: WST Concord context word search.....	132
Figure 11: WST Concord context word search: focus on search horizons.....	134
Figure 12: Incidence of PV use in the LOCNEC corpus expressed in phrasal verb TTR according to the categories of compositionality.....	154
Figure 13: Incidence of PV use in the PLINDSEI corpus expressed in phrasal verb TTR according to the categories of compositionality.....	154
Figure 14: General PV distribution in individual users - PLINDSEI.....	157
Figure 15: General PV distribution in individual users - LOCNEC.....	160

List of abbreviations

AMU	Adam Mickiewicz University
ASCII	American Standard Code for Information Interchange
BAWE	British Academic Written English
BNC	British National Corpus
CAE	Certificate in Advanced English
CANCODE	Cambridge and Nottingham Corpus of Discourse in English
CBA	Corpus-based Approach
CDA	Corpus-driven Approach
CEA	Computer-aided Error Analysis
CECL	Centre for English Corpus Linguistics at the Université Catholique de Louvain in Belgium
CEDEL2	Corpus Escrito del Español como L2
CEEC	Corpus of Early English Correspondence
CEF	Common European Framework of Reference for Languages: Learning, Teaching, Assessment.
CHILDES	Child Language Data Exchange System
CIA	Contrastive Interlanguage Analysis
CL	Corpus Linguistics
CLAWS	Constituent-Likelihood Automatic Word-Tagging System
CLC	computer learner corpus
CLEC	Chinese Learner English Corpus
COCA	Corpus of Contemporary American English
COHA	Corpus of Historical American English
CPE	Certificate of Proficiency in English
CQP	Corpus Query Processor
DAML	DARPA Agent Markup Language Program

DAMSL	Dialog Act Markup in Several Layers
DARPA	Defense Advanced Research Projects Agency
EAP	English for Academic Purposes
EFL	English as a Foreign Language
ELF	English as Lingua Franca
ELFA	English as a Lingua Franca in Academic Settings
ELT	English Language Teaching
ESL	English as a Second Language
ESP	English for Specific Purposes
FLL	Foreign Language Learning
FLT	Foreign Language Teaching
FRIDA	French Interlanguage Database
HKUST	Hong Kong University of Science and Technology Learner Corpus
ICAME	International Computer Archive for Modern and Medieval English
ICE	International Corpus of English
ICLE	International Corpus of Learner English
IFA	Instytut Filologii Angielskiej (School of English)
IL	Interlanguage
IPA	International Phonetic Alphabet
KWIC	Key Word in Context
LINDSEI	Louvain International Database of Spoken English Interlanguage
LLC	London-Lund Corpus
LOB	Lancaster-Oslo-Bergen Corpus
LOCNEC	Louvain Corpus of Native English Conversation
LOCNESS	The Louvain Corpus of Native English Essays
LONGDALE	Longitudinal Database of Learner English
MICASE	Michigan Corpus of Academic Spoken English
MPV	Macmillan Phrasal Verbs Plus
MT	Machine Translation
MWU	Multi-word units
NCP	National Corpus of Polish
NKJP	Narodowy Korpus Języka Polskiego
NL	Native Language
NS	Native Speaker
PL	Polish

PLEC	PELCRA Learner English Corpus
PLINDSEI	Polish LINDSEI subcorpus
POS	Part-of-Speech
PV	Phrasal Verb
SACODEYL	System Aided Compilation and Open Distribution of European Youth Language
SAMPA	Speech Assessment Methods Phonetic Alphabet
SEC	Spoken English Corpus
SLA	Second-language acquisition
SST	Standard Speaking Test
TED	Translanguage English Database
TEI	Text Encoding Initiative
TSM	Tonetic Stress Marks
TTR	Type-Token Ratio
TTS	Text-to-Speech
UCL	University of Louvain
UCREL	University Centre for Computer Corpus Research on Language, Lancaster University
USAS	UCREL Semantic Analysis System
USE	Uppsala Student English Project
VESPA	Varieties of English for Specific Purposes dAtabase
VOICE	The Vienna-Oxford International Corpus of English
WOSLAC	Word Order in Second Language Acquisition Corpora
WST	WordSmith Tools
XML	Extensible Markup Language

Introduction

At the center of the present book lies the phrasal verb, one of the most notorious structures that learners of English have to master. Describing how learners approach the phrasal verb, Celce-Murcia and Larsen-Freeman (1999: 425) state that

most ESL/EFL students will find such verbs strange and difficult. Yet they are ubiquitous in English; no one can speak or understand English (...) without a knowledge of phrasal verbs. Because they don't realize this, some nonnative speakers of English have a tendency to overuse single lexical items where a phrasal verb would be much more appropriate (...)

The reason behind choosing this aspect of investigation for the subject of this book is getting insight into an aspect of English desired from the viewpoint of its native speakers, but problematic when approached by non natives, even at an advanced level of their educational path (e.g. Dagut and Laufer 1985).

After literature review in Chapter 1, where different approaches to phrasal verb definitions are discussed, phrasal verb is defined as a union of “a verb proper and a morphologically invariable particle that function together as a single unit both lexically and syntactically” (Darwin and Gray 1999: 76-77), e.g. *speak up*. In terms of compositionality, there are three possible configurations:

- the verb and the particle retain their lexical, directional meaning (“directional PVs”), e.g. *come back*,
- the verb retains the literal meaning, while the particle adds an aspectual meaning (“aspectual PVs”), e.g. *read through*,
- both the verb and the particle occur with an opaque, idiomatic meaning (“idiomatic PVs”), e.g. *come across*

(Celce-Murcia and Larsen-Freeman 1999; Quirk et al. 1985; Armstrong 2004).

The present book aims to investigate in what ways learners use phrasal verbs, looking into the phenomenon of avoidance. The main research aim is to find out how the semantic categories of PVs are distributed in learners as opposed to native speakers, and whether the distribution of learners' semantic PVs is inversely proportional to the level of PV idiomaticity, as observed in native speakers' use. Namely, whether Polish EFL speakers use the idiomatic semantic category of PVs least of all compositional categories. The major empirical part consists in quantitative and qualitative analysis involving semantic division of phrasal verbs in both learner and native speaker groups into compositionality categories in order to find out proportions of compositionality use to answer the major research questions put forward at the beginning. It needs to be stressed that careful analysis of the English language and meticulous testing procedures (Darwin and Gray 1999) are necessary at this part of this book. The last concern of phrasal verb avoidance is the verification of the influence of L2 exposure on PV use among Polish EFL speakers.

The empirical research was based on a corpus of spoken learner English PLINDSEI (Jendryczka-Wierszycka 2006) and its comparative corpus LOCNEC (Gilquin et al. 2010). Both corpora were POS-tagged in order to gain access to possibly all phrasal verbs in the corpus. The process of improving the accuracy of POS-tagger for the purpose of tagging spoken learner data is the focus of a separate section of the thesis.

this book consists of four chapters. As a springboard for the subsequent semantic analysis of PVs, Chapter 1 discusses different theoretical approaches to the phrasal verb, its classification, and presents studies on phrasal verbs focusing on their use by non-native learners of English. Chapter 2 introduces the methodology crucial for conducting corpus linguistic research and sets it in opposition to the so called armchair linguistic data collection procedure (Fillmore 1992; Rusiecki 2006). Corpus annotation, essential especially for research of linguistics on the verge of vocabulary and grammar, is discussed in section 2.3: *Data Annotation* as an extension of corpus linguistic research. Chapter 3, in turn, provides a thorough description of the application of native language trained annotation tools to non-native spoken corpora on the example of the PLINDSEI corpus, which is the corpus used for the empirical research presented in Chapter 4. The

challenges and ways to address them are also specified therein. The final chapter merges the learner and the semantic research mentioned in section 1.4 into an empirical study, where phrasal verbs used by native and non-native users of English are compared in terms of their general and compositional use. Research questions are put forward and verified in the last chapter and, ultimately, claims concerning the roots of PV avoidance are voiced. They are strictly related to the pedagogy of the phrasal verb, that is how phrasal verbs are taught at schools. As a result, the interpretation of PV avoidance is the manner of PV presentation that has its roots in syntactic and lexical difficulty of PVs. Finally, *Conclusions* sum up the results obtained in the present book. Areas that need further investigation are also indicated. this book comprises 16 consecutively numbered Appendices, which are indispensable for a better comprehension of the present thesis. For reasons of space, they are attached in separate files.