

Thèse de doctorat

de l'Université Sorbonne-Paris Cité
Préparée à l'Université Paris Diderot
Ecole Doctorale 132: Sciences du Langage
CLILLAC-ARP

Discipline : Linguistique Anglaise

**English VP anaphors:
*do it, do this, do that***

Présentée et soutenue publiquement par

Gabriel Flambard

Le 5 décembre 2018

Directeur de Thèse: **Philip Miller**

Composition du jury de soutenance

Anne Abeillé (Présidente)	Université Paris Diderot
Laure Gardelle (rapporteur)	Université Grenoble Alpes
Jonathan Ginzburg	Université Paris Diderot
Andrew Kehler (rapporteur)	University of California at San Diego
Philip Miller (directeur)	Université Paris Diderot
Geoffrey Pullum	University of Edinburgh



Except where otherwise noted, this work is licensed under
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Acknowledgements

I would like to thank my supervisor Philip Miller for his invaluable advice and support throughout the six years it took me to complete this project. The quality of the thesis owes much to his insightful criticism and recommendations for improvement. I am also grateful to Andy Kehler for acting as external supervisor and providing much useful advice at various stages of the project. Thanks also to Anne Abeillé, Laure Gardelle, Jonathan Ginzburg, and Geoffrey Pullum for agreeing to sit on the defence committee alongside Philip and Andy. At the pre-defence meeting last June, Laure and Andy made very helpful suggestions for the final stages of writing, and also brought much-needed reassurance that the thesis could be completed on time and defended by end of year, a prospect which at the time seemed less than certain to me.

This thesis would not have been possible without the doctoral scholarship awarded to me in 2012 by the Université Paris Diderot. The CLILLAC-ARP research team regularly provided funding for travel in France and abroad, giving me many opportunities to present my research at various conferences over the years. More recently, CLILLAC-ARP generously agreed to lend me a MacBook Air after my previous laptop got stolen just months away from the submission deadline. Progress on the thesis would have been much slower without the replacement laptop at my disposal.

The opportunity to spend a year at UC Berkeley as a Graduate Student Instructor in French was certainly the highlight of my career. The position provided funding for a fourth year of doctoral research, and allowed me to gain extensive experience in teaching French as a foreign language. While there, I was fortunate to work with extremely dedicated students who took their language-learning at heart, and to enjoy the unique environment and facilities that Cal has to offer. Financial support for subsequent years came from assistant lecturer positions at the Universities of Tours (2016–2017) and Créteil (since 2017).

Various other people contributed in more punctual but no less important ways. Guillaume Desagulier kindly agreed to share an early version of his book on corpus linguistics with R before its release, enabling me to begin to see the light on many obscure statistical concepts and methods. The corpus analysis presented here, and especially the review of data in Chapter 2, would have been much sketchier and approximate had it not been for this precious resource. Lastly, I would like to thank Saghie Shar-

ifzadeh for getting in touch last summer and sending me her recent paper on VPAs and VPE, which also makes use of data from the BNC. I was not aware of the paper and would most likely have missed it if she had not sent it to me in the first place. While it has not been possible to take all of its findings into account, due to time constraints, the paper brought useful insights for the discussion in Chapters 1 and 2, and will undoubtedly prove an important reference for future research on VPAs.

Abstract

Keywords: Linguistics, Syntax, Discourse Pragmatics, Verbal Anaphora, English, Ellipsis, Proforms.

The present thesis offers a corpus study of the alternation between *do it*, *do this* and *do that* in their use as ‘Verb Phrase anaphors’ (VPAs), in which they refer to a salient action mentioned in previous discourse, typically by means of a VP, or exophorically to a salient action in the speech situation that is not explicitly mentioned in previous discourse. *Do it/this/that* have been little studied in the otherwise extant literature on anaphora and especially VP ellipsis (VPE, e.g., *Kim knows the answer and Pat does too*). This is because it has long been assumed that they are largely interchangeable with each other as well as with *do so* and VPE, so that detailed analysis of their discourse properties was not deemed worth pursuing. The examples below show that this assumption is flawed: in (1), an attested example from the BNC, *do this/that/so* could be used instead of *do it*, but in (3), *do that* is strongly preferred. As for VPE, it is unnatural in (1) and prefers a context of the type in (2).

- (1) They’ve been rescuing companies for so long they do it automatically now, I expect. (AB9, ok: they do this/that/so automatically...)
- (2) They’ve been rescuing companies for so long that whenever they do, it’s always a success.
- (3) He closes his eyes when he speaks and I don’t trust anyone who does that. (AHF, compare: ...anyone who does #this/#it/#so)

Based on a sample of annotated data from the British National corpus (BNC, [Davies 2004](#)), our study will examine the factors driving the alternation between *do it/this/that*. Amongst others, VPA choice is influenced by register, the presence of an adjunct after the VPA, whether or not the antecedent has already been mentioned prior to the antecedent clause, and, to a lesser extent, the saliency of the antecedent and its presumed familiarity to the addressee. *Do it* typically refers to highly salient actions which are then further described by means of an adjunct. *Do this*, by contrast, denotes actions that have not been mentioned before the antecedent

clause, and does not co-occur with adjuncts as often as *do it*. It also allows for much less salient antecedents. *Do that* typically occurs without an adjunct, and sometimes bears much resemblance to VPE in its usage.

Résumé

Les anaphores verbales de l'anglais : *do it*, *do this*, *do that*

Mots-clés : Linguistique, Syntaxe, Pragmatique du Discours, Anaphore Verbale, Anglais, Ellipse, Proformes.

Cette thèse propose une étude de l'alternance entre *do it*, *this* et *do that* dans leur emploi comme anaphores verbales (Verb Phrase anaphors, VPAs), où ils renvoient à une action saillante soit évoquée précédemment dans le discours, le plus souvent via un SV) soit, par exophore, à une action saillante dans la situation discursive mais non évoquée explicitement dans le discours précédent. *Do it/this/that* ont été peu étudiés dans la littérature par ailleurs conséquente sur l'anaphore *and* et en particulier l'ellipse du VP (VP ellipsis, VPE, par exemple *Kim knows the answer and Pat does too*). En effet, on a longtemps considéré que ces trois constructions sont interchangeables entre elles ainsi qu'avec *do so* et l'ellipse, de sorte qu'un examen détaillé de leur propriétés discursives n'a pas été jugé utile. Les exemples ci-dessous montrent que cette supposition est incorrecte : en (1), un exemple attesté tiré du BNC, *do this/that/so* pourraient être employés au lieu de *do it*, mais en (3), *do that* est nettement préféré. S'agissant de l'ellipse, elle est peu naturelle en (1) et préfère un contexte comme celui en (2).

- (1) They've been rescuing companies for so long they do it automatically now, I expect. (AB9, ok : they do this/that/so automatically...)
- (2) They've been rescuing companies for so long that whenever they do, it's always a success.
- (3) He closes his eyes when he speaks and I don't trust anyone who does that. (AHF, compare : ...anyone who does #this/#it/#so)

A partir d'un échantillon d'exemples annotés du British National corpus (BNC, [Davies 2004](#)), notre étude examinera les facteurs qui entrent en jeu dans l'alternance entre *do it/this/that*. Le choix entre les VPAs est déterminé entre autres par le registre, la présence d'un circonstant après l'anaphore, la mention ou non de l'antécédent avant la phrase antécédent, et dans une moindre mesure, la saillance de l'antécédent et la familiarité supposée qu'en a le destinataire. *Do it* renvoie en général à des actions

très saillantes qui sont ensuite décrites plus en détail par le biais d'un circonstant. *Do this*, au contraire, dénote plutôt des actions qui n'ont pas été évoquées avant la phrase antécédent, et son emploi avec un circonstant est moins fréquent que pour *do it*. *Do that* est employé le plus souvent sans circonstant, et son usage présente parfois de grandes similarités avec l'ellipse.

Résumé substantiel

Cette thèse propose une étude de l’alternance entre *do it, this and do that* dans leur emploi comme anaphores verbales (*Verb Phrase anaphors*, VPAs), où ils renvoient à une action saillante soit évoquée précédemment dans le discours (le plus souvent via un SV) soit, par exophore, à une action saillante dans la situation discursive mais non évoquée explicitement dans le discours précédent.

Do it/this/that ont été relativement peu étudiés dans la littérature par ailleurs conséquente sur l’anaphore *and* et plus particulièrement sur l’ellipse du VP (*VP ellipsis*, VPE, par ex. *Kim knows the answer and Pat does too*). Ce phénomène est aussi appelé *Post-Auxiliary Ellipsis* par Sag (1976) ou Miller (2011); Miller and Pullum (2014), terme repris en français par Sharifzadeh (2018) qui parle d’“ellipse post-auxiliaire”. De nombreux auteurs, notamment dans le cadre de travaux sur l’ellipse en syntaxe générative, considèrent en effet *do it/this/that* comme interchangeables entre eux ainsi qu’avec *do so* et l’ellipse dans la plupart des cas. Cette impression, renforcée par l’emploi exclusif d’exemples inventés dans les études de ce type, n’a pas favorisé un examen détaillé des propriétés discursives de *do it/this/that*.

Les rares études sur corpus existantes mettent pourtant en avant des divergences dans l’emploi de *do it/this/that* (Souesme 1985 et plus récemment Sharifzadeh 2018), *do so* (Miller, 2011, 2013) et l’ellipse (Miller, 2011; Miller and Pullum, 2014). Les exemples ci-dessous amènent à nuancer la conception des anaphores verbales évoquée précédemment, en montrant que l’alternance entre n’est pas aussi libre qu’on pourrait le penser. Ainsi, en (1), exemple attesté tiré du British National corpus (BNC, Davies 2004), *do it* peut être remplacé par *do this/that* ou *do so*, mais en (2), également tiré du BNC, *do that* est nettement préféré. Enfin, l’ellipse post-auxiliaire est peu naturelle en (1) (*#They do automatically now...*) et préfère un contexte similaire à celui illustré en (3).

- (1) They’ve been rescuing companies for so long they do it automatically now, I expect. (AB9, ok : *They do it/this/that/so...*)
- (2) He closes his eyes when he speaks and I don’t trust anyone who does that. (AHF, ...anyone who does #this/#it/#so)
- (3) They’ve been rescuing companies for so long that whenever they do, it’s always a success.

L'objectif de la thèse est de rendre compte des variations de ce type sur la base d'un corpus annoté, constitué à partir de données du BNC, afin d'identifier les facteurs d'alternance entre *do it/this/that*. L'analyse repose sur l'hypothèse que les anaphores verbales ont une structure compositionnelle : ce sont constructions transitives de forme *do* verbe lexical + pronom objet (cf. [Simner 2001](#) ; [Stroik 2001](#)). Par conséquent, l'alternance entre les trois formes résulte de la sémantique de *do* et des propriétés discursives des pronoms *it/this/that*, indépendantes de leur emploi dans les anaphores verbales.

La structure de la thèse est la suivante : l'introduction (Chapitre 1) donne un aperçu du problème de l'alternance entre *do it/this/that*, et dresse un état des lieux des études existantes, afin d'isoler les facteurs d'alternance potentiels ou avérés, qui servent de base à l'annotation du corpus.

Le Chapitre 2 détaille la constitution du corpus et le principe d'annotation. Au total, 15 propriétés morphologiques, syntaxiques et sémantiques ont été codées pour chaque exemple. L'analyse statistique du corpus permet d'identifier les facteurs suivants comme étant les plus déterminants pour l'alternance entre anaphores verbales : le registre (*do it* se rencontre plus souvent à l'oral ; *do this* appartient plutôt au registre écrit), la présence d'un circonstant après l'anaphore (*do it* et *do this* sont fréquemment suivis d'un circonstant, mais cela est beaucoup plus rare pour *do that*), et le statut discursif de l'antécédent (*do it* renvoie habituellement à des actions déjà mentionnées avant la phrase antécédent, tandis que *do this* et *do that* renvoient à des actions non mentionnées précédemment).

Le Chapitre 3 propose une analyse compositionnelle de la syntaxe des VPAs, et défend en particulier l'argument selon lequel le pronom seul est anaphorique, et non le SV dans son ensemble.

Le Chapitre 4 revient sur le problème de l'agentivité, souvent avancée comme une contrainte sémantique sur l'antécédent de *do it/this/that*. L'examen du corpus initial et d'un corpus séparé, conçu à partir de requêtes permettant d'obtenir un plus grand nombre d'occurrences non-agentives dans le BNC, montre que les emplois non-agentifs sont attestés mais rares, *do this* étant le plus courant dans cet emploi, alors que *do it* est particulièrement marginal. *Do it/this/that* peuvent exprimer la simple causation par des entités inanimées, ou plus rarement par des animés humains non-volitionnels. Le corpus présente une grande diversité sémantique, mais dans bien des cas la situation décrite reste proche du schéma agentif prototypique impliquant des animés humains. C'est notamment le cas des ac-

tions réalisées par des machines, programmes informatiques, etc., dont l'autonomie d'action facilite l'assimilation avec des agents humains, d'où la désignation collective de "pseudo-agents" sous laquelle ils sont regroupés.

Le Chapitre 5 aborde les cas de discordance syntaxique entre l'anaphore et le "déclencheur d'antécédent" (*antecedent-trigger*), terme par lequel Cornish (1999) désigne un élément qui permet d'accéder à l'antécédent de l'anaphore, permettant ainsi son interprétation. On s'intéresse également aux emplois exophoriques, où l'antécédent n'est pas mentionné explicitement en discours et doit être inféré directement à partir de la situation de discours. Si les VPA ne présentent aucune différence notable dans leur compatibilité avec différents types de discordance syntaxiques, les emplois exophoriques sont au contraire source de variation : *do this* est largement dominant, tandis que *do it* est marginal. Ces divergences sont la conséquence d'une plus grande sensibilité de *do it* à la saillance de l'antécédent : en emploi exophorique, il ne peut renvoyer qu'à des actions se déroulant dans la situation de discours, alors que *do this/that* permettent le renvoi à des actions simplement inférables à partir du contexte extralinguistique.

Le Chapitre 6 examine le rôle de la saillance dans le choix du pronom objet, et montre que ce paramètre joue un rôle limité en-dehors des emplois exophoriques. Bien que *do it* soit le plus fréquemment employé lorsque l'antécédent présente un haut degré de saillance, cet emploi est loin d'être marginal pour *do this/that*, même s'ils permettent de renvoyer à des antécédents moins accessibles, comme expliqué ci-dessus. Par ailleurs, *do this/that* sont parfois attestés dans des exemples où l'antécédent est hautement saillant, mais *do it* est bien moins acceptable.

Le Chapitre 7 analyse l'alternance entre *do it/this/that* sous l'angle de la fonction discursive de la phrase anaphorique : description des propriétés de l'action au moyen d'un circonstant, identification de l'agent, etc. 12 fonctions possibles sont distinguées, mais certains exemples échappent à la classification établie et relèvent d'autres fonctions encore mal définies.

Les fonctions les plus couramment rencontrées sont la description de propriétés supplémentaires de l'action (*adjunct-choice*, "choix du circonstant"), l'emploi d'une anaphore comme complément, le plus souvent d'un verbe et en particulier d'un auxiliaire modal, ainsi que, dans une moindre mesure, certaines fonctions typiques de l'ellipse, telles que le choix d'une alternative polaire (*polarity-choice*, "choix de la polarité") ou l'identification de l'agent (*agent-choice*, "choix de l'agent").

Cependant, on observe des divergences importantes dans les fonctions typiques de chaque forme. La description de propriétés supplémentaire est le plus souvent introduite par *do it* ou éventuellement *do this*, mais très rarement par *do that*. À l'inverse, *do that* est préféré à *do it/this* dans les cas de “choix de polarité”, ce qui montre son aptitude à remplir des fonctions semblables ou identiques à celles de l'ellipse. Enfin, on observe une moindre fréquence de *do this* lorsque la phrase anaphorique est limitée à un jugement modal sur l'action.

Contents

1	Introduction	1
1.1	VP anaphora	1
1.2	State of the art on VPA	6
1.2.1	Overview	6
1.2.2	Theoretical accounts	7
1.2.3	Corpus studies	9
1.2.4	Experimental studies	12
1.3	Structure of the thesis	12
2	Corpus coding: methods and results	14
2.1	The corpus: data collection methods	14
2.1.1	Overview	14
2.1.2	Searching for VPAs in the BNC	15
2.2	Annotation scheme: features coded	22
2.2.1	General features of the text	22
2.2.2	Morphological features of the VPA	23
2.2.3	Adjuncts	23
2.2.4	Denotation the antecedent and VPA clauses	23
2.2.5	Structure of the antecedent-trigger	24
2.2.6	Structure of the VPA	24
2.2.7	Discourse status of the antecedent	25
2.2.8	Agentivity	25
2.2.9	Function of the VPA clause	25
2.3	Overview of corpus data	25
2.3.1	Register	25
2.3.2	Finiteness of main verb <i>do</i>	30
2.3.3	Adjuncts	32
2.3.3.1	A note on adjuncts and finiteness	36

2.3.4	Adjunct Types	36
2.3.5	Identity of subjects	40
2.3.6	Identity between states of affairs	42
2.3.7	Structure of the antecedent-trigger	44
2.3.7.1	Complements	47
2.3.7.2	Auxiliaries	47
2.3.7.3	Anaphoric constructions	47
2.3.7.4	Non-parallel antecedent-triggers	48
2.3.7.5	Other non-finite structures	49
2.3.7.6	No antecedent-trigger	49
2.3.8	Clause type of the antecedent-trigger	51
2.3.9	Main or subordinate status of the antecedent clause	52
2.3.10	Structure of the VPA	53
2.3.11	Syntactic type of the VPA clause	54
2.3.12	Main or subordinate status of the VPA clause	56
2.3.13	Discourse status of the antecedent	57
2.3.14	Agentivity	59
2.4	Conclusion	61
3	Compositionality	63
3.1	The compositionality problem	63
3.1.1	The internal structure of VP anaphors	63
3.1.2	Literature review	65
3.2	Evidence for compositionality	71
3.2.1	<i>It</i> as pro-VP	71
3.2.1.1	Ambiguity revisited	71
3.2.1.2	Properties of the VPA/NPA readings	73
3.2.1.3	Evidence for the pro-VP status of <i>it</i>	75
3.2.1.4	Choice of pro-form: <i>do it</i> and <i>do this/that</i>	78
3.2.1.5	<i>Do it</i> and <i>do this/that</i> : discourse properties	79
3.2.1.6	Main verb status of <i>do</i>	80
3.3	Apparent problems for the compositional analysis	81
3.3.1	* <i>Do</i> + VP and the antecedent of <i>it</i>	82
3.3.2	Antecedent and antecedent-trigger	83
3.4	Conclusion	85

4	Agentivity	89
4.1	Introduction	89
4.2	Agentivity constraints on VPA antecedents	90
4.3	Defining agentivity	97
4.3.1	Agents and non-agents	97
4.3.2	Linguistic tests for agentivity	98
4.3.3	Semantic features of agentivity	99
4.4	Non-agentive VPAs: a corpus study	103
4.4.1	Overview of the BNC data	103
4.4.2	Data collection methods	106
4.5	Results: frequency of non-agentive VPAs	108
4.5.1	Method 1: pronominal subjects	108
4.5.2	Method 2: STEM texts	110
4.5.3	Overall results	112
4.5.4	Influence of finiteness	113
4.5.5	Influence of register	115
4.5.6	Summary	115
4.5.7	Agentivity and adjuncts	116
4.5.7.1	Adjuncts	116
4.5.7.2	Adjuncts and finiteness with non-agentive VPAs	118
4.6	Typology of non-agentive VPAs	120
4.6.1	Causality and animacy	121
4.6.2	Semantic roles coding	121
4.6.3	Results of the semantic role annotation	124
4.6.4	Causes	125
4.6.5	Instruments	128
4.6.6	Themes	130
4.6.7	Experiencers	133
4.6.8	Patients	133
4.6.9	Stative antecedents	134
4.6.10	A note non-on on agentive <i>do so</i>	136
4.6.11	Acceptability of non-agentive VPAs	137
4.7	Conclusion	139
5	Antecedent mismatch	142
5.1	Introduction: Mismatch with VPA and VPE	142
5.1.1	Deep and surface anaphora	145

5.1.2	Experimental accounts of mismatch with VPE and <i>do it</i>	147
5.1.3	Summary of findings and notes on materials	152
5.2	Antecedent mismatch with VPAs	157
5.2.1	Mismatch and inference	158
5.2.1.1	Mismatch and agentivity	159
5.2.2	Passives	161
5.2.2.1	Presence or absence of a passive agent	161
5.2.2.2	Alternation with VPE	165
5.2.3	Nominal antecedent-triggers	167
5.2.3.1	Conditions on nominal antecedent-triggers	168
5.2.3.2	Overview of corpus data	169
5.2.3.3	Deverbal nouns	170
5.2.3.4	Nominal gerunds	173
5.2.3.5	Other nominal antecedent-triggers	173
5.3	Exophora and inferred antecedents	174
5.3.1	Exophoric VPAs	177
5.3.2	Discourse inference	181
5.4	Conclusion	186
6	Saliency and pronoun choice in VP anaphora	188
6.1	Introduction	188
6.2	Choice between <i>it/this/that</i> outside of VPA	189
6.2.1	Properties of <i>it</i>	189
6.2.2	Choice between <i>this/that</i>	190
6.2.3	Constraints on the choice between <i>it</i> and demonstratives	193
6.3	<i>It/this/that</i> in VPA	196
6.3.1	<i>Do it</i> : high saliency/old information	196
6.3.2	<i>Do this/that</i> : speaker-inclusion/exclusion and related features	199
6.3.3	Topic continuation and importance of the referent	202
6.4	On the saliency of actions and events	203
6.4.1	Linguistic factors	204
6.4.2	Extralinguistic factors	205
6.4.3	Familiarity beyond discourse	207
6.5	Discourse status of the antecedent	209
6.5.1	Discourse-old antecedents with <i>do this</i> and <i>do that</i>	212
6.5.1.1	<i>Do this</i>	213

6.5.1.2	<i>Do that</i>	214
6.5.2	Conclusion: discourse-newness and VPAs	216
6.6	Saliency	216
6.6.1	Corpus study: applying the GH to VP anaphors	217
6.6.1.1	Corpus design and annotation scheme	217
6.6.1.2	Results and discussion	218
6.6.2	Alternative coding: saliency levels	221
6.6.3	Cognitive statuses and saliency	222
6.6.4	Conclusion	226
7	Function of the VPA clause	227
7.1	Introduction	227
7.1.1	Background	228
7.1.2	VPA clause functions	230
7.1.3	Data: frequency of VPA functions	233
7.2	Adjunct-choice	235
7.2.1	Coding scheme	235
7.2.2	Data	237
7.3	Different occurrence of action	239
7.3.1	Coding scheme	239
7.3.2	Data	239
7.4	Polarity-choice	240
7.4.1	Coding scheme	240
7.4.2	Data	243
7.4.2.1	Factors preventing VPE	244
7.4.2.2	Alternation with VPE	246
7.5	Agent-choice	249
7.5.1	Coding scheme	249
7.5.2	Data	252
7.5.2.1	Agent-choice with modality contrasts	254
7.5.2.2	Agent-choice and polarity contrasts	257
7.6	Modal-choice	258
7.6.1	Coding scheme	258
7.6.2	Data	260
7.6.2.1	Frequency of modals with VPAs	260
7.6.2.2	Modal forms and meanings	261
7.7	Other complements	264
7.7.1	Coding scheme	264

7.7.2	Data	266
7.8	Directive	269
7.8.1	Coding scheme	269
7.8.2	Data	270
7.9	Intended action	270
7.9.1	Coding scheme	270
7.9.2	Data	271
7.10	Temporal location	273
7.10.1	Coding scheme	273
7.10.2	Data	274
7.11	Adjuncts	276
7.11.1	Coding scheme	276
7.11.2	Data	276
7.12	VPA is sentence subject	280
7.12.1	Coding scheme	280
7.12.2	Data	280
7.13	Other functions	282
7.14	Conclusion	286
8	Conclusion	288
8.1	Summary of findings	288
8.2	Perspectives for future research	291
8.3	Pending issues in VPA alternation	291
8.3.1	Problems for the compositional analysis	291
8.3.2	Experimental work	292
8.4	Comparison with other anaphoric expressions	292

Chapter 1

Introduction

1.1 VP anaphora

The present thesis offers a corpus study of the alternation between the set of *do* + pronoun constructions *do it*, *do this* and *do that* based on a sample of data from the British National Corpus (BNC, [Davies 2004](#)). Specifically, we will be concerned with the use of these constructions as ‘Verb Phrase anaphors’ (henceforth VPAs) in which they refer to a salient action either mentioned in previous discourse, typically by means of a VP, or exophorically to an action that is salient in the speech situation but not explicitly mentioned in previous discourse. Attested examples of VPA *do it/this/that* from the BNC are given below; those in (1a)–(1c) are anaphoric (or more precisely endophoric), while those in (2a)–(2c) are exophoric.¹

- (1) a. They’ve been rescuing companies for so long they do it automatically now, I expect. (AB9)
- b. John yawned, and scratched his stomach. While he was doing

1. In order to facilitate the reading of the examples, which are sometimes a bit long because broader context is relevant to the discussion, both the antecedent (more specifically the antecedent-trigger, as discussed below) and the relevant instance of VPA are underlined. For exophoric uses, only the VPA is underlined.

CHAPTER 1. INTRODUCTION

this Maggie took a pan of boiling water from the cooker and laid it by the side of the sink. This was John's shaving water. (AN7)

- c. Parliament did not give the Secretary of State the power to fix poll tax levels. He's told everybody he was fixing poll tax levels, and he hadn't got the power to do that. (KRT)
- (2) a. Masha, kneeling, was pressing her mouth to Rozanov's, attempting to force life into his body. At last I got through, demanded an ambulance, and had to ask Masha for the number of the room. She pulled back from him, sitting on her haunches. 'It's no good. I'm not even sure if I'm doing it right, or if it's the right thing to do.' (AEO)
- b. Rodney cracked two eggs into the frying pan. 'I'm not doing this every morning,' he said. 'With so many of us we ought to set a rota for cooking.' (AOR)
- c. Somewhere quite close a pheasant called. Philip imitated it. The boy stared at him. 'How d'you do that, then?' he said. (ABX)

In their use as VPAs (at least when they refer endophorically to the action, as in (1)), *do it/this/that* alternate with *do so*—itself a VPA—and, under certain conditions, VP ellipsis (VPE, also called 'Post-Auxiliary Ellipsis' by Sag 1976; Miller 2011). Below are constructed variants of the examples in (1) with *do so* substituted for *do it/this/that*.

- (3) a. They've been rescuing companies for so long they do so automatically now, I expect. (= 1a)
- b. John yawned, and scratched his stomach. While he was doing so... (= 1b)
- c. He's told everybody he was fixing poll tax levels, and he hadn't got the power to do so. (= 1c)

VPE, as [Miller \(2011\)](#) showed, does not in fact alternate quite so easily with *do it/this/that* or *do so*, as it is subject to different discourse constraints. For instance, it is dispreferred if followed by a non-contrastive adjunct, as in (1a). It is also normally excluded in non-finite uses such as (1c). The following variant of (1a) provides a more suitable context for VPE:

- (4) They've been rescuing companies for so long that whenever they do, it's always a success.

VPE, and to a lesser extent *do so*, have both been relatively well-studied. VPE has been the object of much research in the generative literature ([Sag, 1976](#); [Sag and Hankamer, 1984](#); [Dalrymple et al., 1991](#); [Kehler, 2002](#); [Merchant, 2004](#); [Kertz, 2010, 2013](#), amongst many others) as well as in psycholinguistics ([Murphy, 1985](#); [Tanenhaus and Carlson, 1990](#); [Arregui et al., 2006](#); [Kertz, 2010, 2013](#); [Phillips and Parker, 2014](#)). While *do so* has not received as much attention as VPE, it has still been the focus of a number of studies, including a thesis by [Houser \(2010\)](#) (see also [Bouton 1970](#); [Lakoff and Ross 1976](#); [Michiels 1978](#); [Sobin 2008](#); [Miller 2013](#)). Arguably, there is still much left to explain about VPE and *do so*, but their usage is much better-understood than that of *do it/this/that*.

Studies of *do it/this/that* are far and few between, though they are discussed briefly in descriptive grammars of English such as [Quirk et al. \(1985\)](#) or [Huddleston and Pullum \(2002\)](#), and sometimes get mentioned in passing in the context of more general discussion of VPE (e.g., [Hankamer and Sag, 1976](#); [Merchant, 2004](#)) and other types of anaphor or ellipsis ([Culicover and Jackendoff, 2005](#)). A number of experimental studies, informed by [Hankamer and Sag \(1976\)](#)'s proposed distinction between 'deep' and 'surface' anaphors (we will review them in Chapter 5), have compared the behaviour of *do it* and VPE under the specific condition of 'antecedent mismatch' with a view to determining whether or not VPE requires a syntactically parallel antecedent, as suggested by [Hankamer and](#)

Sag. Studies of this type are, for the most part, chiefly concerned with testing a syntactic hypothesis about VPE, rather than explaining the discourse conditions determining its usage and alternation with *do it/this/that* or *do so*. *Do it* is in fact taken as representative of the *do* + pronoun set of VPAs, and *do this/that* are always left out of consideration. This treatment stems from the widely-held assumption in the literature on VPE that *do it/this/that* as well as *do so* are syntactically interchangeable and largely identical from the point of view of discourse, and present virtually no relevant semantic difference with VPE other than the requirement that the antecedent of *do it/this/that* should be non-stative, as shown by the example in (5)

- (5) a. Kim knows the answer and Pat does too.
b. Kim knows the answer and Pat #does it/this/that too.

Miller and Hemforth (2014, p. 3) sum up on the usual line of thinking on VPAs as follows:

Generative studies of PAE and verbal anaphors such as *do so* or *do it/this/that* have never addressed the discourse conditions on these constructions. In fact, it has been tacitly assumed that there is no difference in their uses beyond those resulting from some version of the deep/surface distinction of Hankamer and Sag (1976) and the fact that *do* in *do so* and *do it/this/that* is the main verb.

The premise of our study is that this tacit assumption is incorrect, and that even though the choice between *do it/this/that* is at first glance relatively free, the very possibility of choosing between *it/this/that* as the object pronoun is good reason to think that there are relevant differences in usage between them, and that these differences are largely attributable to the properties of the object pronoun, either *it* or demonstrative *this/that*. Further, it may be argued that treating *do it/this/that* as (near-)equivalents

of one another makes as little sense as analysing *it/this/that* in a similar fashion.

The examples in (1a) support the idea that VPA choice is relatively free, since in each of them *do it/this/that* could all be replaced with another VPA. However, it is often the case that *do it/this/that* are not interchangeable and that only one or two of the three is felicitous. Before reviewing empirical evidence of this in detail, it is worth bearing in mind some essential discrepancies in the use of VPAs, VPE and *do so*.

As argued by Miller (2011); Miller and Hemforth (2014); Miller and Pullum (2014), *do it/this/that* are dispreferred in contexts typical of VPE, such as replies to polar questions, or when confirming or denying a previous assertion (6).

- (6) He shops in women's.
B: No, he doesn't. [COCA]
[Compare #No, he doesn't do it /this /that.] (Miller and Pullum, 2014, ex. 11, p. 9)

Do so is more similar to *do it/this/that* in its discourse properties, but Miller (2011) points out that it is especially typical of formal written register (for instance, it is frequent in academic writing) and also prefers to denote the same specific situation (state of affairs) as the antecedent sentence. Compare the variant of (1b) in (7), with identical states of affairs in both clauses, to that in (8) where *do so* denotes a different situation from the antecedent clause.

- (7) John yawned, and scratched his stomach. While he was doing so Maggie took a pan of boiling water from the cooker...
- (8) John yawned, and scratched his stomach. #He did so all the time.

As just explained, *do it/this/that* do not always freely alternate. (9a) is an example where *do that* is preferred over *do this* or *do it* (9b). However,

in the constructed variant in (9c), they are both acceptable alongside *do that*.

- (9) a. He closes his eyes when he speaks and I don't trust anyone who does that. (AHF)
b. ...and I don't trust anyone who #does this/it.
c. ...and I hate when he does it/this/that.

A central aim of the thesis will be to account for such differences on the basis of the semantics of the object pronoun. Yet even in 'free choice' examples, where VPAs appear to be more or less equivalent, one may still wonder what led to the choice of a particular form when others could have been used instead. Thus, in addition to factors of alternation strongly favouring one of *do it/this/that*, we will also consider what motivates their usage in contexts where there is little or no difference in acceptability between them.

1.2 State of the art on VPA

1.2.1 Overview

While it is fair to say that VPA alternation has long been an under-studied topic in comparison to VPE and *do so*, it has not been entirely unexplored either. Putting aside studies like [Hankamer and Sag \(1976\)](#) and others in a similar vein, whose primary focus is VPE rather than VPA (these will be discussed in Chapter 5), existing accounts of VPA may be divided into purely theoretical ones, corpus-based analyses, and experimental ones. The first kind are found in descriptive grammars of English such as [Quirk et al. \(1985\)](#) or [Huddleston and Pullum \(2002\)](#), and tend to be limited to a handful of invented examples. Corpus-based studies vary in size and

scope. [Souesme \(1985\)](#) is an early example of the sort of task undertaken here—an extensive analysis of VPAs based on attested examples. Unlike ours, his study also encompasses VPE, *do so* and other anaphoric *do*-constructions such as *do something*, *do the same*, etc. [Miller \(2011\)](#), based on a small sample from the COCA corpus, shows how the discourse conditions of VPE differ from those of VPAs generally, and makes some proposals on the choice between *do so* and *do* + pronoun VPAs, and between *do it/this/that* themselves. [Miller and Pullum \(2014\)](#) are chiefly concerned with exophoric (‘antecedentless’) uses of VPE, but highlight a number of important discrepancies in the use of VPE/VPA in both exophoric and endophoric contexts. More recently, [Sharifzadeh \(2018\)](#) expands on the proposals of [Miller \(2011\)](#) and goes further into detail regarding the alternation between *do it/this/that*. Experimental studies are also quite different in scope: [Simner \(2001\)](#) and [Simner et al. \(2003\)](#) focus on the processing of *do it* depending on whether it is a VPA or NPA; [Miller and Hemforth \(2014\)](#) provide an experimental investigation of [Miller \(2011\)](#)’s proposed discourse constraints on VPE. The next sections give a short survey of these earlier studies, highlighting their respective contributions to the study of VPA and how these will shape the work carried out here. Discussion of their findings will be kept to a minimum as they will be reviewed in more detailed in Chapter 2, since the annotation scheme used in coding the corpus is largely informed by earlier proposals.

1.2.2 Theoretical accounts

[Quirk et al. \(1985, p. 877ff\)](#) refer to VPAs as ‘complex pro-forms which are substitutes for predicates or predications’. They explain that *do it/that* ‘combine the substitute function of *do* and the coreferential function of *it/that*’ (presumably this is also true of *do this*, which [Quirk et al.](#) do not mention) and differ syntactically from *do so* in allowing passivisation (e.g., *It/this/that/*so was done*).

Quirk et al. (1985) also point out that *that* in *do that* can and often does carry prosodic stress, whereas *it/so* in *do it/do so* are always unstressed. They also suggest that *do it* resists non-agentive antecedents at least for some speakers, and *do that* or *do so* are preferred (it is unclear whether or not they regard *do this* as also allowing non-agentive uses).

Lastly, Quirk et al. point to information status as another factor of choice between VPAs. They suggest that *do that* tends to introduce information that is ‘treated to some extent as new or contrastive’ while *do it* conveys information that is ‘entirely given’ (p. 877). The following examples are proposed to illustrate this contrast (small capitals indicate stress):

- (10) Is Connie still trying to light the stove? She should have DONE it by now.
- (11) Are you trying to light the stove with a match? I wouldn’t do THAT.

In the first example, *still* implies that the addressee has previous knowledge of the action (it is ‘entirely given’, say Quirk et al.), while in the second the antecedent is probably new information. The contrast between new/given information is of particular interest here since it is applicable to the choice between *it* and demonstratives independently of VPAs (see e.g., Fillmore, 1982; Gundel, 1993; Strauss, 2002).

Huddleston and Pullum (2002) provide a very similar account of VPAs, although they do mention *do this* as well. They consider *do it/this/that* as regular transitive constructions, citing evidence from passivisation and the possibility of prosodic stress on the demonstrative pronoun.

The agentivity constraint on the antecedent is claimed to hold for all three of *do it/this/that*, in contrast to *do so*, and is considered a true requirement (rather than a usage preference as with Quirk et al. 1985). *Do it* is also claimed to preferentially refer to the same action as its antecedent, and this preference is extended to *do that*. Huddleston and Pullum also mention exophoric uses; for instance, in *Don’t do that!*, the pronoun could

refer to the action being performed by the addressee, and likewise *He was doing this* could be uttered ‘as I demonstrate the action in question’ (p. 1532). Also mentioned is the possibility of remnant PP[*to*] complements, as in (12):

- (12) They questioned Jill for over an hour before letting her go: I hope they don’t do that to me. (ex. 51, p. 1533)

Information status is also advanced as a factor of choice between VPAs, at least for *do it* and *do that*; for instance, it is proposed that an answer with *do it* in (13) is likely if A’s resignation is already known to the addressee, while *do that* would be more likely if it is new.

- (13) A: I’ve sent in my resignation.
B: Why did you do it?
B: Why did you do that?
(ex. 54, p. 1534)

Again, it is not specified how *do this* behaves in this respect, but presumably it is associated with new rather than given information, as [Huddleston and Pullum](#)’s example below would suggest:

- (14) We need to make absolutely clear what the goals of the various courses are : only if we do this will people be able to make an informed choice between them. (ex. 48, p. 1532)

1.2.3 Corpus studies

[Souesme \(1985\)](#) is one of the few extensive corpus studies of VPAs. He draws attention to various important properties of VPAs, such as their non-idiomatic status; the preference for agentive antecedents, the common use of adjuncts after VPA, saliency as a condition on exophoric uses, and the influence of the proximal/distal contrast on *do this* and *do that*. Souesme is unsystematic in his analysis, however, as a number of the parameters

discussed for *do it* are not further examined with *do this/that*, and his conclusions on the factors licensing VPA alternation are not always clear.

Drawing on data from the Corpus of Contemporary American English (COCA, [Davies 2008](#)), [Miller \(2011\)](#) suggests that the choice between *do it/this/that*, *do so* and VPE is influenced by, amongst others, register, the presence of an adjunct following the VPA, whether the VPA sentence refers to the same state of affairs as the antecedent sentence and has the same subject, as well as the saliency of the antecedent. All of these features have been included in the corpus annotation and are therefore discussed further in the next chapter. As for saliency, [Miller \(2011\)](#) argues that *do it* does not require as salient an antecedent as *do so* or VPE, which has the highest requirement for the saliency of its antecedent. He observes that *do this/that* appear to have a lower requirement than *do it*, and further that *do this/that* may in some cases be dispreferred if the antecedent is salient enough for use of *do it*, as in (15) where the first occurrence *do this* makes the antecedent salient, leading to a preference for *do it* over *do this*.

- (15) I didn't know I couldn't do this...so I just did it/#this (ex. 11d, edited)

Another central claim in [Miller's](#) study is that VPE is subject to specific discourse conditions that differ from those of VPA, and is much less acceptable in contexts where these conditions do not obtain. For instance, VPE is favoured when choosing a branch of a polar alternative (see (6) above), but it is dispreferred if followed by an adjunct describing additional properties of the antecedent action (16).

- (16) Using 19th century technology they not only accomplished it but did it so well that it was later incorporated into the London subway system and remains in use in the 21st century. [#but did so well that...] (11a, p. 89)

Conversely, as already explained, *do it/this/that* are unnatural in typical

VPE contexts (see (6)).

Miller and Pullum (2014) provide a more exhaustive statement of the discourse conditions on VPE, and show that the dispreference for VPAs in contexts that satisfy the conditions for VPE extends to exophoric uses, as shown by the example in (17a) and the variant in (17b):

- (17) a. [Entering a construction site, A hands a helmet to B]. B: Do I have to? (24b, p. 16)
b. [Same context] B: #Do I have to do it/this/that?

Sharifzadeh (2018) examines the choice between VPAs and VPE on the basis of an annotated sample of data from the BNC. Due to the paper being very recent, it has not been possible to take all of its results into account here. We will comment briefly on some relevant observations made in the paper regarding *do it/this/that*. Sharifzadeh makes a distinction between VPA and NPA, as we do here (see above and Chapter 3 for further discussion of this issue) and also distinguishes idiomatic uses of *do it* such as (18) from others.

- (18) You really have that winning attitude. You know you can do it...get out there and have it all! (CDK, cited p.4 as ex. 5 and abridged here)

Her corpus is annotated with the following parameters: whether the example is of VPA or NPA; the type of reference (anaphoric, cataphoric, or exophoric); register (spoken or written), agentivity of the antecedent, and aspect (telic or atelic). Since all of these features except for aspect are coded in our corpus, discussion of her results will be carried out in Chapter 2.

We now give a brief survey of the data on reference type and aspect. Sharifzadeh observes that 48% of *do it* tokens are of VPA, as compared to 67% and 68% for *do this/that* respectively. *Do it* is much more common than the demonstrative variants in NPA, and less common in exophoric

uses (9%, as compared to 15% and 16% for *do this/that*).

As regards aspect, the antecedent is in most cases telic, but *do this* appears to be more frequent than *do it/that* with atelic antecedents (25.9%, as opposed to 15.4% with *do that* and 17.7% with *do it*).

1.2.4 Experimental studies

The alternation between *do it/this/that* has, to my knowledge, never been studied experimentally. Existing experimental studies always consider *do it* in relation to VPE, and do not discuss how their results might extend to *do this/that*. The analysis carried out here will be limited to a corpus-based approach, but it should provide fertile ground for experimental testing in future research.

1.3 Structure of the thesis

The rest of the thesis will be organised as follows: Chapter 2 explains the design of the corpus and the annotation scheme, giving an initial overview of the statistical data for each feature coded. Chapter 3 clarifies the syntax of VPAs and makes a case for treating them as ordinary transitive constructions in which the pronoun alone, rather than the entire VP, is anaphoric. Chapter 4 reexamines the question of whether VPAs require an agentive antecedent, as several authors have claimed. Our data support the weaker version of this claim which views agentive predicates only as the preferred denotation of VPAs; *do it/this/that* may express simple non-agentive causation by animate or inanimate entities, non-causal processes, and in some rare cases states. Chapter 5 concentrates on ‘antecedent mismatch’ with VPA. Though VPAs do not appear to differ significantly with respect to their allowance of mismatches, we will see that the discourse context in which mismatched VPAs occur is quite different from that of attested

1.3. STRUCTURE OF THE THESIS

mismatches with VPE. Chapter 6 explores the role of the saliency of the antecedent, a factor which has been shown to affect the choice between *it/this/that* independently of VPA. Chapter 7 considers VPA alternation from the angle of the structure in which the VPA occurs and the discourse function of the VPA clause.

Chapter 2

Corpus coding: methods and results

2.1 The corpus: data collection methods

2.1.1 Overview

The analysis relies primarily on attested data from the BNC (Davies, 2004), a corpus of British English comprised of both written and spoken speech from the 1980s-90s.¹ In this respect, our data differ from those of Miller (2011, 2013); Miller and Pullum (2014), which focused on American English with data from the COCA. We will also include examples from the COCA where relevant to illustrate a particular aspect of VPA usage, but the BNC remains the primary source of data, and the only one for the central corpus used throughout this thesis.

The sample used here is comprised of 900 occurrences—500 of *do it* and 200 each of *do this* and *do that*—taken in equal proportion from the written and spoken discourse. Written data consist exclusively of fiction

1. A more exhaustive description of the corpus, including details of the design, markup conventions and the nature of the texts included, is available online at www.natcorp.ox.ac.uk/docs/URG/.

2.1. THE CORPUS: DATA COLLECTION METHODS

works from the 1980s-1990s, while the spoken data include, for the most part, radio or TV broadcasts, as well as (in a very small number of cases) recorded conversations or meetings. The motivations for the choice of the fiction and broadcast sections will be discussed further below.

The size of the corpus, as well as the ratio of written to spoken speech, is entirely arbitrary: these parameters were determined before collecting the data, and the corpus as a whole is obviously not representative of the respective frequency of VPAs in discourse, nor of their distribution across registers. The decision to have an arbitrary sample size is due to the impossibility of searching directly for VPAs in the BNC—as explained further below, it is only possible to search for any combination of *do* followed by *it* or *this/that*, not to specify that the two must form a constituent.

The smaller size of the *do this/that* sample, on the other hand, is simply linked the genesis of this project, which started life as a study on *do it* only, and was later extended to *do this* and *do that* when already at an advanced stage. Due to the amount of work involved in collecting and annotating the data, it was not possible to include as many occurrences of *do this/do that* as would have been desirable. To avoid skewing the data further, it was decided to have the same number of occurrences for *do this* and *do that*, and an equal proportion of spoken and written speech.

Each occurrence in the corpus was annotated with a range of features either identified as parameters of VPA choice in previous studies, or considered as such on the basis of preliminary empirical evidence. The full set of features and the coding scheme for each of them are discussed in [2.2](#); the remainder of the introduction being devoted to the methods used to collect examples of VPAs despite the limitations of the BNC.

2.1.2 Searching for VPAs in the BNC

All examples were obtained using the online search interface at corpus.byu.edu/bnc and supplemented with full-text data from the Oxford

CHAPTER 2. CORPUS CODING: METHODS AND RESULTS

Text Archive (ota.ox.ac.uk/desc/2554). These data greatly facilitated the comprehension of a large number of examples where the interpretation of the VPA was not immediately clear from the limited context provided by the online interface.

The query used to retrieve occurrences of *do it/this/that* is the following:

```
[do] it|this|that
```

which searches for all forms of *do* followed by either *it*, *this* or *that*. Unfortunately, as explained, the query does no more than that: the search interface offers no way to restrict results to occurrences where the *do* + pronoun combination is a VP rather than some other non-constituent sequence. This is because the BNC is tagged only at the part-of-speech level (including punctuation marks) and does not encode features above word level, such as phrasal category (NP, VP...) or sentence boundaries. Since there is also no encoding of grammatical function, there is no way to ensure that the pronoun is the object of the *do*, for instance to rule out examples of inversion between auxiliary *do* and *it/this/that* as the subject pronoun. The BNC also lacks any form of semantic annotation, hence any indication of anaphoric relations, which is a serious limitation for the task at hand, especially when examining syntactic or semantic properties of the antecedent clause, such as mismatch or the agentivity of the antecedent. Finally, the part-of-speech tagging itself is not always as fine-grained as would be desired. Crucially, it does not differentiate between main verbs and auxiliaries, which would have made it easier to separate VPA from SAI cases. Thus, a significant amount of noise had to be filtered out from the initial data when building the corpus. Common examples of irrelevant patterns are shown below (with the *do* + pronoun sequence underlined for clarity); these include:

SAI in interrogatives or tags, as well as certain negative structures:

2.1. THE CORPUS: DATA COLLECTION METHODS

- (19) How long does it go on for? (D97)
- (20) There we go what does this actually say (F8U)
- (21) How long di- how long ago did that happen? (FY6)
- (22) Doesn't look very steady though does it! (D97)
- (23) [N]owhere in any of their literature does it actually say what time of day the seminar is. (F8U)

VPE with auxiliary *do* followed by a subject pronoun in a different sentence (this is fairly limited and typically occurs in an *if*-clause, with the pronoun as the subject of the following main clause).

- (24) 'Yes, of course, it is flattering but I couldn't let it rule my life and I won't.'
'If you did it would drive you crazy.' (CH1)

And, more generally, any occurrences where the verb and pronoun together do not form a constituent.

With *do this/that*, we find examples with determiners rather than pronouns. While the BNC does, in theory, have a separate tag for both, it turns out that *this/that* are in fact never classified as pronouns, always as determiners.

- (25) [W]ell I mean if, if he's gonna do this seven page thing then [...] (D97)
- (26) You cannot do that sort of thing if you're a priest. (CH6)
- (27) How about a statement of what we intend to do this week [...] (D97)

With *do that*, there are some examples of relative clauses, though they are uncommon (28):

- (28) After this they have little to do that is essential until the models are proved and the experimental drawings must be modified and converted into production drawings. (HNV)

Results for *do it* also include cases of NP anaphora (NPA), where the pronoun is simply anaphoric for a previously-occurring direct object, as the underlining in (29) indicates:

(29) They do whatever is necessary. And they do it quickly. (CH2)

Do it is also attested in certain idiomatic uses often expressing some notion of ‘success’, as in the last two occurrences of (30), or simply ‘causation’, where the meaning is roughly that the *do it* action is sufficient to bring about some other situation (31)–(32).

(30) Football coach, Bear Bryant, is also a farmhand and in the old days farmhands used to guide a team of animals in front of the plough, encouraging them and making sure they all pulled together. But he also coaches a football team and his secret for teamwork is this. What he says is that if anything goes badly, he tells the team, ‘I did it’. If anything goes reasonably well, he says, ‘We did it!’ But when everything goes really well, he says to them, ‘You did it!’ (FSN)

(31) Anyway, I got the part and it certainly couldn’t have been my charm or personality that did it [=got me the part]. (HRF)

(32) ‘And if you are too tired to write you must come to see me in any case, and then I shall know you are safe. I have a good mind to ask you to serve full-time on my reporting staff, but I am selfish enough to want you to go on writing for me these exposes of low life, particularly the way in which such misery afflicts women. You are doing your sex a service, you know,’ he finished. That did it! That just did it! Sally-Anne could have jumped up and down for joy. (HGE)

Idioms like the above will not be considered further, since they differ in significant ways from VP-anaphoric uses of *do it*. First, the interpretation

2.1. THE CORPUS: DATA COLLECTION METHODS

is more or less fixed, rather than obtained via the processing of a preceding segment of discourse where the action is described in full (an ‘antecedent-trigger’, as Cornish (1999) calls it). Secondly, the action described is often non-agentive, i.e. non-volitional, and the subject is commonly inanimate, whereas in anaphoric uses, *do it* (like *do this/that*) primarily denote agentive actions and thus occurs with animate subjects. Lastly, *do this/that*, *do so* or VPE are all excluded here since they would not yield the relevant interpretation.² Idiomatic uses raise a number of independent problems which go somewhat beyond the scope of a thesis on VPA alternation. In particular, it would be interesting to establish the exact range of idiomatic meanings that can be distinguished, and why *do it* alone would be suitable to express them. It is also an open question whether idiomatic meanings bear any relation to anaphoric ones. All of these questions are left open for further research, our goal here being to examine the alternation between *do it/this/that* in their anaphoric uses only.

Also of note are ‘taboo avoidance’ uses in which *do it* serves as a euphemism for actions the speaker prefers to avoid mentioning, e.g., sexual intercourse, or anything scatological (33):

- (33) It happens in the blink of an eye. You rush out the door to drop your kids off at school, or to run an errand, only to return and find that your pet has ‘done it’. (carpetcleanerseattle.com/services/pet-odor-stain-elimination)

Taboo uses seem to be halfway between idioms and regular VP anaphora: there is usually no antecedent-trigger present, and the meaning is largely predictable, but it is still context-dependent in a way that the idioms above are not. It is also largely a matter of common sense or world knowledge, rather than actual lexical understanding of the expression as with true idioms. In (33), the unpleasant ‘discovery’ made by the owner

2. (31) is an ambiguous case between the type illustrated in (32) and VPA. Both interpretations seem possible, either ‘cause success’, or VPA, as seen by the fact that replacement by *do this/that* is not entirely excluded, contrary to (32).

on returning home is typically that the pet has relieved themselves on the floor, and the and owner has to clean up; this is a matter of stereotypical knowledge. Out of context, simply uttering, for instance, *My cat/dog has done it* would not so easily be understood.

Aside from the irrelevant cases, problems often arise with the comprehension of the VPA examples themselves. A recurring problem in the spoken portion of the corpus is that the recordings contain many inaudible passages marked only as ‘unclear’ in the transcripts. This usually makes the VPA difficult or impossible to interpret, and many examples had to be left out since they proved too difficult to understand without the missing data. This problem was especially acute in the ‘conversation’ and ‘meeting’ sections of the BNC, where it is not uncommon for an entire stretch of conversation to be entirely incomprehensible. For this reason, the ‘broadcast’ section, which contained fewer incomprehensible examples, was chosen over these two sections as the primary source of spoken data. An unclear example from a transcribed conversation section is shown in (34) as it appears in the BNC (with codes for different speakers in parentheses). The unclear fragments, as well as the lack of information as to the context in which the conversation takes place, prevents successful interpretation of *do it*.

- (34) (SP:PS01F) Lie down. (SP:PS01B) Snuggled up on that blanket, little un. (unclear) (laugh) (SP:PS01G) Lie down (unclear). (SP:PS01F) You’d do it for for Bert. (KB1)

The fiction section, from which the written data were taken, is itself not exempt from lack of clarity. In that case, though, the problem stems not from the example being incomplete, but from the need to have read a substantial portion of the book in order to fully understand the passage in which the VPA occurs. One such example appears in (35):

- (35) Today is your lucky day. Grey is your lucky colour. You are my lucky star. And give it time! says Sebastian. Don’t wash it for

2.1. THE CORPUS: DATA COLLECTION METHODS

three days! Give it time, I know you, you're so impatient. Yes, Sebastian, no Sebastian, my cup runneth over. Now stop that. Shut up. Whatever happens will. No buts. Just get on. Shop for food. Easy, do it in my sleep. Do it. Shut the hell up. Time's moving on. (AOL)

Having explained the difficulties in retrieving relevant examples of VPAs, we must now say a word about the specific BNC sections chosen to build the corpus. As already mentioned, the written component is entirely made up of fiction texts, while the spoken component is comprised of a variety of spoken texts, mostly broadcasts (84% of the data), with the remainder coming from various other spoken texts. The fiction sample was restricted to prose (the BNC also includes drama and poetry), while the spoken texts also include interviews, meetings and lectures. These additional sections were used since not enough relevant or understandable examples were found in the broadcast section. The spoken sections used are listed below, with the abbreviated name used in the BNC labelling, and a short description of the contents.

- S_brdcast_documentary TV documentaries (*Private Eye, The Money Programme...*)
- S_brdcast_discussn Television discussions on a variety of topics (crime, mental health, transport, etc.)
- S_brdcast_news Radio or TV news
- S_interview_oral_history Interviews conducted as part of the Nottingham Oral History Project
- S_interview Job interviews, as well as some radio interviews
- S_lect_nat_science Lectures in natural science delivered at British universities

- `S_lect_polit_law_edu` Lectures in politics, law and related topics delivered at British universities
- `S_lect_soc_science` Lectures in social science delivered at British universities.
- `S_meeting` Meetings between members of various organisations

2.2 Annotation scheme: features coded

Overview

The annotation was carried out in Excel, with a column for the VPA and its extended context, and an array of formal or semantic features of the example: whether main verb *do* is finite or non-finite, whether the VPA is followed by an adjunct (and if so, what kind of adjunct it is), whether the antecedent is agentive or non-agentive, and so on. The full set of features³ is listed below with a brief explanation of each. The coding scheme and results of the annotation are discussed further in the following sections.

2.2.1 General features of the text

Medium The medium of production of the text, coded as written or spoken accordingly.

Section : The BNC section from which the example is taken, for instance `W_fict_prose` for fiction texts, `S_brdcast_news` for news broadcasts, and so on.

3. The annotation also includes features that are purely informational and of no linguistic relevance, such as the text in which the example occurs or the number of the example.

Discourse Type Whether the VPA occurs in a passage of *written* or *spoken* discourse, the latter including fictional dialogues.

2.2.2 Morphological features of the VPA

Finiteness Whether main verb *do* is *finite* or *non-finite*.

Verb form The particular form of *do* occurring in the example, for instance *1SG_past* (first-person singular, past) for *I did it*, *3PL_pres* (third-person plural, present) for *They do it*.

2.2.3 Adjuncts

Adjunct : Whether or not the VPA is followed by an adjunct, coded respectively as *Adj* or *NoAdj*.

Adjunct Type The semantic type of the adjunct (if there is one), for instance *manner*, *means*, *location*, etc. Reflexives of the type in *I did it myself* are also considered as adjuncts and are coded as *reflexive*.

2.2.4 Denotation the antecedent and VPA clauses

There are two features of interest here: whether the antecedent and VPA clauses have coreferent subjects, and whether or not they denote the same specific action or ‘state of affairs’ (SoA, in the sense of token-identity, rather than type-identity). In our coding, these features are called ‘subject identity’ and ‘SoA identity’ for short.

Subject Identity is coded as *same* if the two clauses have coreferent subjects, and *different* if they are not.

SoA Identity is likewise coded as *same* or *different* according to whether the antecedent and VPA clauses denote the same SoA.

Additionally, NA is used wherever identity between subjects or SoAs cannot be determined, for instance in exophoric uses.

2.2.5 Structure of the antecedent-trigger

Structure The structure containing the antecedent-trigger. Categories distinguished include, amongst others, `main VP` for finite VP predicates, `Vcomp` for verb complements, `passive VP`, `NP`, `VPA` and `VPE`. NA is used if there is no antecedent-trigger.

Details Further details of the structure if applicable. For instance, the label `Vcomp_try` means that the VP is the complement of *try*.

Antecedent clause type The syntactic type of the antecedent clause, such as `declarative` or `interrogative`.

Antecedent clause subordination Whether the antecedent clause is a `main` or `subordinate` clause.

2.2.6 Structure of the VPA

Structure The structure containing the VPA was coded using the same categories as with the antecedent-trigger except for those that are not applicable, such as `VPE` or `NP`. Although VPAs can be passivised, only active examples were collected in the corpus, so that the `passive_VP` label was never used.

Details Further details of the VPA structure, coded with the same categories as the antecedent-trigger.

VPA clause type The syntactic type of the VPA clause.

VPA clause subordination The `main` or `subordinate` status of the VPA clause.

2.2.7 Discourse status of the antecedent

The antecedent is considered discourse-new if it has not been mentioned before the antecedent clause, and discourse-old if it has. The coding uses the label `new` and `old` to capture this distinction.

2.2.8 Agentivity

Whether the antecedent is agentive or non-agentive. Agentive to us means ‘volitional’; we do not consider non-volitional animate and inanimate causes as agents (though they are regarded as such by some other authors).

2.2.9 Function of the VPA clause

By ‘function’ of the VPA clause is meant the purpose of making reference to the antecedent action with a VP anaphor, such as describing further properties of the action (e.g., when, why or how it was done; we call this usage ‘adjunct-choice’), stating whether the action was or was not performed (‘polarity-choice’), amongst others. Because we distinguish many such functions, each requiring explanation of the annotation criteria prior to analysis of the data, discussion of this feature deserves a chapter unto itself, and is postponed to Chapter 7.

2.3 Overview of corpus data

2.3.1 Register

Previous work

[Miller \(2011\)](#) points out that VPE is extremely common in spoken discourse, while *do so* is typical of formal written discourse (notably aca-

demic academic writing). He does not comment on the distribution of *do it/this/that* across registers, and at present no such data are available outside of the present study.⁴

Coding scheme

Recall that as far as the medium is concerned, the ratio of spoken to written speech is exactly 50%, a deliberate choice made in designing the corpus. While this means we do not have data on the distribution of VPAs across registers in general, we can still examine the influence of register if we take fictional dialogues into account when counting the frequency of occurrences in spoken speech. Of course, the frequencies observed in dialogue may not necessarily be representative of the general situation with respect to spoken discourse, but they are nevertheless a good indicator.

Results

Table 2.1 gives an overview of the distribution of VPAs across registers in our corpus. Since we are including fictional dialogues in the spoken portion, VPAs in spoken discourse (real or imagined) are much more frequent. However, *do it* is more frequent in spoken discourse than *do this* or *do that*.

The χ^2 -test shows a highly significant association between VPA choice and register ($p < 0.001$), suggesting that the choice of VPA varies between written and spoken discourse. To further examine the influence of register on VPAs, we look at Pearson residuals, a measure of deviation from the

4. In her corpus study of VPAs based on BNC data, Sharifzadeh (2018) concludes that *do it* is the most frequent in spoken discourse, followed by *do that*, *do this* and *do so*. In written discourse, *do so* and *do it* are about as frequent, but they occur much more often than *do this/that*. However, her data are not precise enough since her search results, obtained using the BNC tag [vd*] (equivalent to [do]) followed by a pronoun or *so*, include just the sort of irrelevant occurrences mentioned earlier, such as SAI and various other non-constituent sequences.

2.3. OVERVIEW OF CORPUS DATA

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Spoken	402 (80.4%)	130 (65%)	145 (72.5%)	677
Written	98 (19.6%)	70 (35%)	55 (27.5%)	223
TOTAL	500	200	200	900

Table 2.1: Frequency of VPAs in written and spoken discourse (including fiction dialogues) in the BNC

expected frequency. [Desagulier \(2017, p. 184\)](#) presents them as a more efficient method than manually comparing the frequencies of different observations, and explains how they are to be interpreted:

If a Pearson residual is positive, then the corresponding observed frequency is greater than its expected frequency. If a Pearson residual is negative, then the corresponding observed frequency is smaller than its expected frequency. Second, the more the Pearson residual deviates from 0, the stronger that effect.

[Desagulier](#) explains that a convenient way to visualise residuals graphically is by means of a Cohen-Friendly association plot ([Cohen, 1980; Friendly, 2000](#)). According to [Desagulier](#), in an association plot,

Each cell of data is represented by a tile whose height is proportional to the corresponding Pearson residual and whose width is proportional to the square root of the expected counts. The area of the box is therefore proportional to the difference between observed and expected frequencies. The dotted line is the baseline. It represents independence. If the observed frequency of a cell is greater than its expected frequency, the tile appears above the baseline and is shaded black. If the observed frequency of a cell is smaller than its expected frequency, the tile appears below the baseline and is shaded red.

In what follows we will make use of the variant of the association plot developed by Zeileis et al. (2007) as part of the `vcd` package for R, which Desagulier cites as an easier way to visualise Pearson residuals. This plot, created by the function `assoc()`, uses selective shading depending on the value of the residual: only tiles with a residual smaller than 2 or greater than -2 are shaded; in blue if the frequency is more than expected, and in red if the frequency is less than expected. The opacity of the shading also varies with the residual, a darker shading indicating a stronger deviation from the expected frequency.

Thus, Fig. 2.1 shows that *do it* is less frequent than would be expected in writing (as indicated by the red shading of the cell) whereas *do this* is more frequent in writing than expected (as indicated by the blue shading).

Details of the data for the fiction sample only are shown in Table 2.2. *Do it* is much more frequent in dialogue, whereas *do this* occurs primarily in narration. *Do that* shows a small preference for narration but is more evenly distributed between both registers.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Dialogue	152 (60.8%)	30 (30%)	45 (45%)	227
Narration	98 (39.2%)	70 (70%)	55 (55%)	223
TOTAL	250	100	100	450

Table 2.2: Frequency of VPAs in narration and dialogue in BNC fiction texts

The association between VPA choice and register is again highly significant ($p < 0.001$). The association plot also in Fig. 2.2 shows comparable results to those for the overall sample in Fig. 2.1, with a strong association preference for *do this* over *do it*. Here, though, we also observe a symmetrical tendency in dialogue, with *do it* being more frequent than expected, and *do this* less frequent. In both the overall data and the fiction sample only, *do that* seems to be largely unaffected by register.

2.3. OVERVIEW OF CORPUS DATA

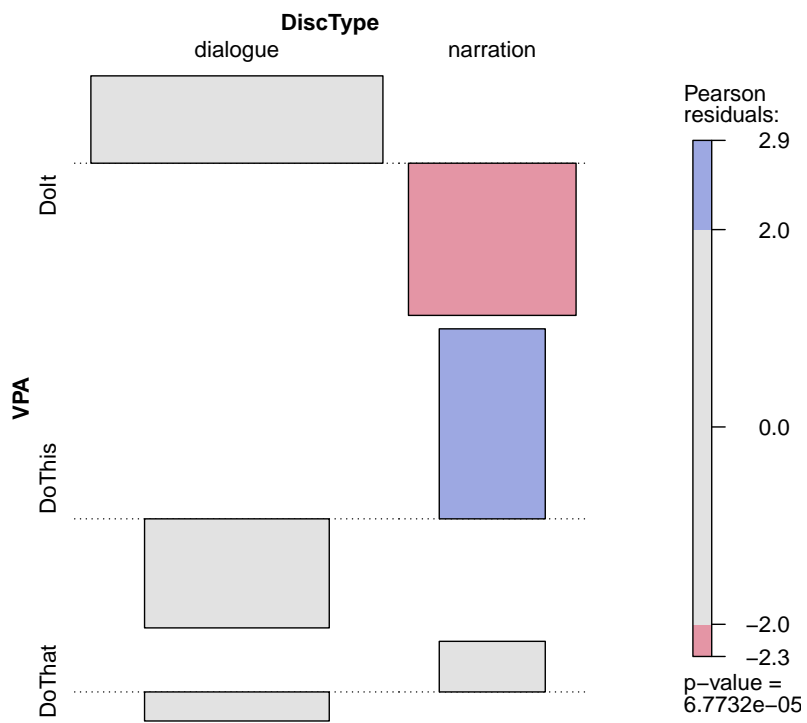


Figure 2.1: Cohen-Friendly association plot showing Pearson residuals for VPAs and register (“Disc[ourse] Type”)

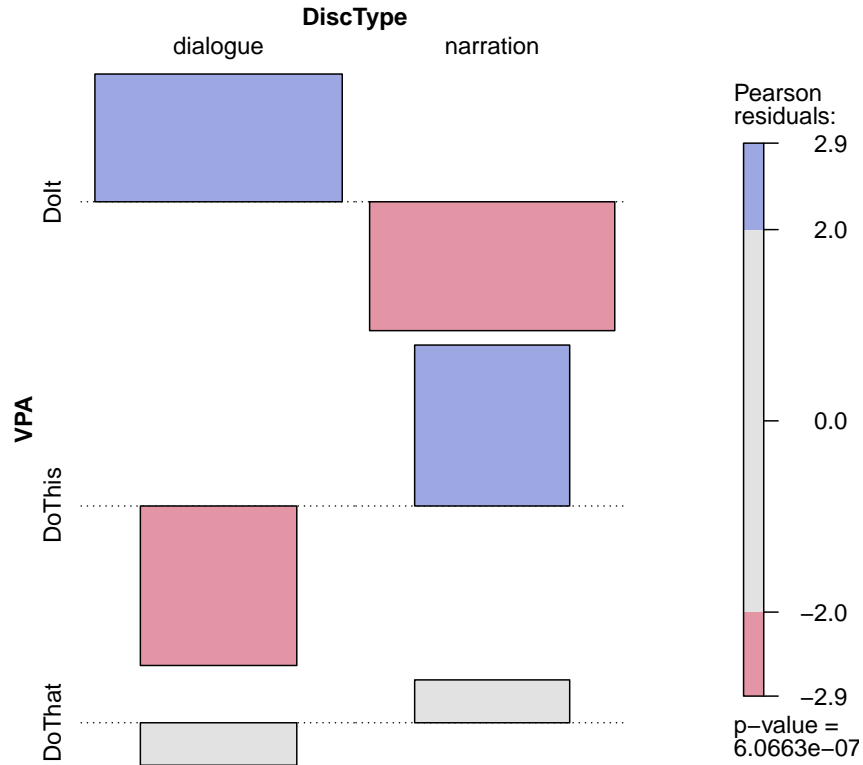


Figure 2.2: Association plot showing Pearson residuals for the distribution of VPAs in narration and and dialogue in the fiction sample

2.3.2 Finiteness of main verb *do*

Previous work

Miller (2011) observed that non-finite forms of VPAs are much more frequent than finite forms. The finite vs. non-finite distinction is important as it determines the possibility of VPE with auxiliary *do*, which is normally only finite, except in some variants of British English. The following is one such example of non-finite VPE with *do* in British English (often called ‘British *do*’):

- (36) “Pity you didn’t hear what Malfoy’s actually doing, though.”
 “I couldn’t have done, could I? That was the whole point, he was refusing to tell Snape.”
 (J.K. Rowling, *Harry Potter and the Half-Blood Prince*)

There is some divergence between authors as to whether this should be treated as a case of VPE with auxiliary *do* or a separate construction with non-auxiliary *do* (see e.g., Miller (2002); Méry (2002); Aelbrecht (2010) for different positions and discussion).

Houser (2010) shows that the impossibility of non-finite VPE in most dialects also affects *do so*, which has a higher ratio of stative antecedents if it is non-finite, precisely because VPE—which would be expected when referring to states—cannot occur. Miller (2013) provides further experimental evidence of this effect. However, Miller (2011) does not comment on the influence, if any, of finiteness on VPA choice beyond its implications for alternation with VPE.

Coding scheme

The annotation encodes both whether main verb *do* is finite or non-finite, as well as the specific form of *do*. With finite forms, the corpus contains details of the number, person and tense (in the same column). For instance, *I did it* is coded as 1SG_past (short for ‘first-person plural, past tense’) and *They do it* as 3PL_pres (‘third-person plural, present tense’). For non-finite forms, we distinguish between bare infinitives (V-inf) and *to*-infinitives. Non-standard *done* used as the past form of *do* (e.g., %*I done it*=*I did it*) is considered finite.

Results

Table 2.3 shows the frequency of finite and non-finite forms with VPAs. Our data replicate Miller’s earlier observations based on COCA data: non-

finite forms are also predominant in the BNC sample examined. Clearly, though, the finiteness of *do* has no influence on VPA choice, as all occur in finite and non-finite uses about as often. The χ^2 -test shows no significant association between finiteness and VPA choice ($p = 0.8$).

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Finite	97 (19.4%)	42 (21%)	39 (19.5%)	178 (20%)
Non-finite	403 (80.6%)	158 (79%)	161 (80.5%)	722 (80%)
TOTAL	500	200	200	900

Table 2.3: Frequency of finite and non-finite VPAs

As regards the frequency of different verb forms (see Table 2.4), the infinitive is dominant, with the base form being the most frequent. The χ^2 -test is not applicable due to the frequency of some forms being lower than 5, so that it is impossible to assess the significance of the verb form for the choice between VPAs. From the data in Table 2.4 *do this* appears to be more frequent than *do it* and *do that* in the *V-ing* form, and less frequent in the base form. Analysis of Pearson residuals (the difference between expected and observed frequency) shows a strong deviation for these two forms with *do this*: 3.2 with *V-ing* (much more than expected), and -3.1 with the bare infinitive (much less than expected).

Though it seems unlikely that VPA choice is strongly influenced by the form of *do* based on the above data, the observations made for *do this* are at least noteworthy. Further statistical analysis would be required to assess the significance of these results.

2.3.3 Adjuncts

Previous work

The use of adjuncts after VPAs has been noted in various earlier studies. Souesme (1985) observes that *do it* is commonly followed by an adjunct

2.3. OVERVIEW OF CORPUS DATA

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
V-inf	176 (35.2%)	37 (18.5%)	64 (32%)	277 (30.8%)
TO-inf	103 (20.6%)	53 (26.5%)	50 (25%)	206 (22.9%)
V-ing	67 (13.4%)	51 (25.5%)	29 (14.5%)	147 (16.3%)
past-ppl	60 (12%)	17 (8.5%)	18 (9%)	95 (10.6%)
3SG-past	30 (6%)	19 (9.5%)	10 (5%)	59 (6.6%)
1PL-past	4 (0.8%)	4 (2%)	1 (0.5%)	9 (1%)
1PL-pres	6 (1.2%)	3 (1.5%)	3 (1.5%)	12 (1.3%)
3PL-past	8 (1.6%)	4 (2%)	6 (3%)	18 (2%)
3PL-pres	20 (4%)	4 (2%)	6 (3%)	30 (3.3%)
3SG-pres	4 (0.8%)	6 (3%)	5 (2.5%)	15 (1.7%)
1SG-past	4 (0.8%)	1 (0.5%)	1 (0.5%)	6 (0.7%)
1SG-pres	6 (1.2%)	1 (0.5%)	4 (2%)	11 (1.2%)
2SG-past	5 (1%)	0	1 (0.5%)	6 (0.7%)
2SG-pres	7 (1.4%)	0	2 (1%)	9 (1%)
TOTAL	500	200	200	900

Table 2.4: Distribution of possible verb forms for main verb *do* in VPAs. Abbreviations: *inf* = infinitive, *ppl* = participle, *pres* = present, SG = singular, PL = plural; numbers 1–3 indicate person.

but does not give a figure, and does not specify whether adjuncts are also common with *do this/that*. Miller (2011) reports that 60% of finite *do it* occurrences in his corpus have an adjunct, as compared to 83% for finite *do so*. Again, no data for *do this/that* are provided. Miller also shows that VPE is dispreferred with adjuncts (at least non-contrastive adjuncts, which describe some additional property of the antecedent action), making them an important factor in the choice between VPAs and VPE. Lastly, Miller (2013) points out that with finite VPAs, using an adjunct can prevent the VPA sentence from being redundant by merely repeating the antecedent clause. For instance, removing the adjunct in (37) makes the question nonsensical, since the relative clause presupposes that the women in question carried out abortions:

(37) Do you think that the ladies who did abortions did it mainly for

money [?] (FXX)

It must be stressed, however that this cannot be the sole motivation for using an adjunct. It is not necessarily the case that removing the adjunct will result in redundancy; this problem will only arise in specific contexts, mainly when the antecedent and VPA clause denote the same state of affairs, as is the case in (37).

Coding scheme

The Adj label is only applied if the adjunct follows the VPA; adjuncts occurring anywhere else in the sentence are not taken into account, and the example is labelled NoAdj. Moreover, the annotation does not indicate the number of adjuncts; Adj is used if there is at least one (a rapid survey of the data suggests, however, that in most cases there is only one adjunct).

Results

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Adjunct	243 (48.6%)	85 (42.5%)	40 (20%)	368 (40.9%)
No Adjunct	257 (51.4%)	115 (57.5%)	160 (80%)	532 (59.1%)
TOTAL	500	200	200	900

Table 2.5: Frequency of adjuncts after VPAs in the sample

We observe a slight predominance of the No Adj case overall⁵, but there are notable discrepancies between VPAs. A higher frequency of adjuncts is observed with *do it/do this*; *do that* very strongly prefers to occur without a subsequent adjunct. The χ^2 -test shows a highly significant association

5. Because there are no general data on the relative frequency of sentences with or without an adjunct following the VP, it is impossible to determine whether the presence of an adjunct is more typical of sentences with a VPA as VP than of sentences in general.

2.3. OVERVIEW OF CORPUS DATA

between VPA choice and the presence or absence of a subsequent adjunct ($p < 0.001$).

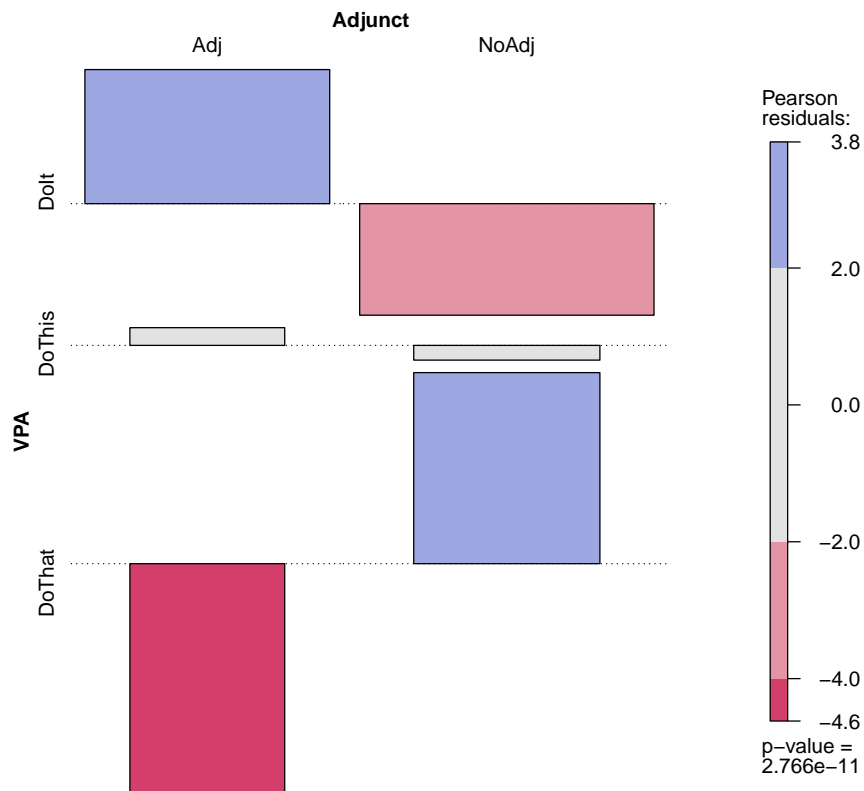


Figure 2.3: Association plot showing Pearson residuals for the frequency of adjuncts after VPAs

The association plot in Fig. 2.3 shows that the pattern *do it* + adjunct is more frequent than expected, whereas *do it* – adjunct is less frequent than expected. With *do that*, the frequency of adjuncts is much lower than expected, and the frequency of occurrences without as subsequent adjunct is somewhat higher than expected. *Do this* appears to occur indifferently with or without an adjunct after the VPA.

2.3.3.1 A note on adjuncts and finiteness

Given that VPE is dispreferred if there is an adjunct, and excluded (except for cases of so-called ‘British *do*’, which some authors classify as instances of VPE) if *do* is non-finite, then we would expect adjuncts to occur more often with finite VPAs, where VPE is in competition, than with non-finite VPAs where it is not. The data in Table 2.6 match the expected distribution, as the frequency of adjuncts is higher in finite than non-finite cases. The χ^2 -test shows a highly significant association between finiteness of main verb *do* and the presence or absence of an adjunct ($p < 0.001$), suggesting that alternation with VPE plays a role in the use of adjuncts after *do it/this/that*. However, there is an asymmetry—the preference for having an adjunct after finite VPAs is much smaller than the preference for not having one after non-finite VPAs.

	Finite	Non-Finite
Adjunct	96 (53.9%)	272 (37.7%)
No Adjunct	82 (46.1%)	450 (62.3%)
Total	178	722

Table 2.6: Frequency of adjuncts after VPAs with finite and nonfinite *do*

Despite the small difference in frequency between occurrences with and without a subsequent adjunct in finite cases, analysis of the Pearson residuals (see Fig. 2.4) lend support to our initial hypothesis: +Adj cases are more frequent than expected, whereas –Adj cases are less frequent.

2.3.4 Adjunct Types

Previous work

Miller (2011, p. 91) comments briefly on the types of adjuncts attested with *do so* in his corpus, saying that manner and means adjuncts are the

2.3. OVERVIEW OF CORPUS DATA

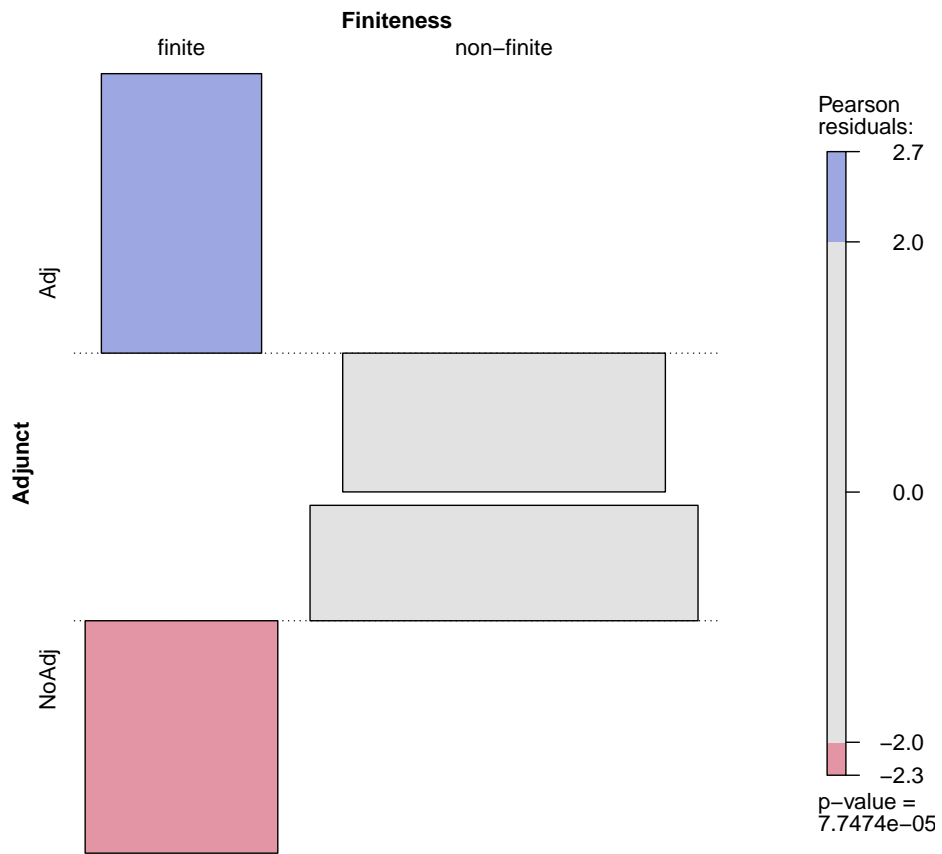


Figure 2.4: Association plot for the frequency of adjuncts after finite and non-finite VPAs

most frequent, but he does not discuss the types of adjuncts encountered with *do it/this/that*. Other studies mentioning the use of adjuncts after VPAs do not go into the details of which semantic categories are attested.

Coding scheme

Our classification is largely inspired from that proposed by [Huddleston and Pullum \(2002, p. 665-6\)](#), but deviates from it in various ways. For some categories, we simply use a different name; for instance, time adjuncts such as *They did that in 1972* (HV2) are called ‘temporal location’ adjuncts by [Huddleston and Pullum](#). For others, we omitted distinctions between certain closely-related similar categories, such as ‘instrument’ adjuncts which we consider a subcategory of means adjuncts.

Results

Table 2.7 shows the distribution of different adjunct types with *do it*, *do this* and *do that*. Time, manner and means adjuncts are the most frequent overall, and a further fifteen less common categories are attested, such as frequency, duration, purpose etc. Since no data are available on the frequency of different adjunct types of in English overall, it is impossible to decide whether the above frequencies simply reflect the general situation with respect to adjuncts, or whether VPAs pattern differently with adjuncts than other non-anaphoric VPs. It seems unlikely, however, that location adjuncts occur as infrequently as they do here, or that they are less frequent than manner or means adjuncts, given that spatial location is such a fundamental type of denotation.

Due to the large number of categories, some of which occur only once in the sample, the χ^2 -test cannot be used to assess whether VPA choice is affected by the semantic type of the following adjunct. We will comment briefly on the types of adjuncts most often encountered after each VPA.

2.3. OVERVIEW OF CORPUS DATA

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Time	63 (25.9%)	11 (13.1%)	11 (28.2%)	85 (23%)
Manner	64 (26.3%)	12 (14.3%)	2 (5.1%)	78 (21%)
Means	24 (9.9%)	15 (17.9%)	1 (2.6%)	40 (11%)
Remnant	10 (4.1%)	11 (13.1%)	8 (20.5%)	29 (8%)
Frequency	8 (3.3%)	13 (15.5%)	3 (7.7%)	24 (7%)
Beneficiary	15 (6.2%)	4 (4.8%)	2 (5.1%)	21 (6%)
Reflexive	14 (5.8%)	3 (3.6%)	1 (2.6%)	18 (5%)
Location	9 (3.7%)	4 (4.8%)	5 (12.8%)	18 (5%)
Purpose	12 (4.9%)	4 (4.8%)	0	16 (4%)
Duration	4 (1.6%)	5 (6%)	3 (7.7%)	12 (3%)
Cause	7 (2.9%)	1 (1.2%)	0	8 (2%)
Condition	5 (2.1%)	0	0	5 (1%)
Comitative	2 (0.8%)	0	2 (5.1%)	4 (1%)
Negative	2 (0.8%)	0	1 (2.6%)	3 (1%)
Concessive	1 (0.4%)	1 (1.2%)	0	2 (1%)
Degree	1 (0.4%)	0	0	1
Result	1 (0.4%)	0	0	1
Aspectual	1 (0.4%)	0	0	1
TOTAL	243	84	39	366

Table 2.7: Semantic types of adjuncts occurring after VPAs

Manner is the most frequent type with *do it* and *do that*, while *do this* is most often followed by a means adjunct. Pearson residuals also suggest a close association between *do this* and frequency adjuncts (3.19), as well as between *do that* and location and remnant adjuncts (2.8 and 2.4 respectively). On the other hand, *do that* is especially infrequent with manner adjuncts (−2.1), and *do it* with remnant adjuncts.

Furthermore, there appears to be some variation in how VPAs pattern with adjuncts of the same category. For instance, *do this* is commonly encountered with means adjuncts of the form *by + VP[ing]* (38), which are less typical with *do it/that*.

(38) We try to make sure the judge has all the facts before arriving at a

decision. Mostly we do this by ensuring that our client gets to talk to the decision-takers. (AB9)

2.3.5 Identity of subjects

The coreference of subjects between the antecedent and VPA clauses was identified by Miller (2011) as a strong usage preference for *do so*, which preferentially denotes the same state of affairs as its antecedent-trigger (98% of cases in his corpus have coreferential subjects) and thus to have the same subject (otherwise the VPA sentence would denote a different occurrence of the same type of action). Miller does not specify how these parameters play into the use of *do it/this/that*, but reports that 46% of finite *do it* tokens have coreferent subjects in the antecedent and VPA clause (no data on *do this/that* are provided).

Coding scheme

As long as the two clauses have overt subjects, the annotation is usually straightforward; we use the label `same` when subjects are coreferent across clauses and `different` when they are not. Some difficulties arise with null subjects, passive VPs (especially agentless ones) and nominal antecedent-triggers; in all such cases the coding depended on whether the null subject of the antecedent clause was interpreted as coreferent with that of the target clause, even if it was also a null subject. Consider for instance (39): the referent of *him* in *Let him do it* is the speaker's son, and the pronoun is therefore coreferent with the implicit subject of the imperative.

- (39) He ran to the bureau. He pulled out the bottom drawer, grabbed all the leaflets from the estate agents, all the newspaper cuttings of houses for sale, and threw them on the floor. They scattered everywhere [...]

2.3. OVERVIEW OF CORPUS DATA

'Pick them up, there's a good boy,' said his Mum, picking up the cutting nearest to her, a half page of photos of houses for sale.

'Leave it,' said his Dad. 'Let him do it.' (ABX)

With passive VPs, the relevant criterion is, of course, the identity of the passive agent, not the subject of the passive sentence. If the agent is expressed, coreference is determined in the same way as for active sentences, by checking whether the passive agent is coreferent with the subject of the VPA clause. In the case of interrogatives with *who* as the subject, the interrogative pronoun is taken to be coreferent with an the agent of the sentence if it is unknown (40):

- (40) It was clear that funds were being embezzled but who could be doing it? Could it possibly be the accountant Edward Morris? (CKD)

Nominal antecedent-triggers may have a syntactic subject, but this is actually rare in the corpus; the only example of this type appears in (41), where the VPA sentence can be read as *It was time for them to do that (= make love)*, and therefore has the same subject as the antecedent clause.

- (41) And bit by bit their lovemaking turned into a dry ritual which caused Rita no actual pain, only a lingering, grey regret. Eventually she found other things to do when, in his view, it was time to do that. (AOR)

Elsewhere, establishing coreference between subjects is again a matter of determining the interpretation of null subjects, as with VP antecedent-triggers.

NA cases, where identity of subjects remains uncertain, are those where the antecedent-trigger is unavailable, namely exophoric uses, those where the antecedent is inferred from previous discourse rather than through a specific linguistic structure, passages of quoted speech that do not include the antecedent-trigger, and occurrences where the antecedent-trigger is

too distant to be considered as the primary source for the interpretation of the VPA.

Results

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Same	335 (67%)	123 (61.5%)	132 (66%)	590 (65.6%)
Different	118 (23.6%)	26 (13%)	48 (24%)	192 (21.3%)
NA	47 (9.4%)	51 (25.5%)	20 (10%)	118 (13.1%)
TOTAL	500	200	200	900

Table 2.8: Frequency of VPAs having the same subject as the antecedent-trigger, or a different one

Table 2.8 shows the frequency of VPAs with coreferent or non-coreferent subjects as well as cases where coreference cannot be determined. We observe a preference for coreferent subjects overall, with mostly little variation across VPAs except for *do this* which is more often attested than *do it/do this* in contexts where identity between subjects cannot be determined. This is due to its frequent use in exophora.

The χ^2 -test shows a highly significant association between VPA choice and subject identity ($p < 0.001$), but this is mainly due to the influence of *do this*. As shown in Fig. 2.5, the frequency of NA cases with *do this* is much higher than expected, while the frequency of non-coreferent subjects is much lower than expected. Beyond this, however, there is no evidence that subject identity is a decisive factor in VPA alternation.

2.3.6 Identity between states of affairs

Previous work

Quirk et al. (1985) as well as Huddleston and Pullum (2002) have also commented on the role of identity between SoAs, though unlike Miller

2.3. OVERVIEW OF CORPUS DATA

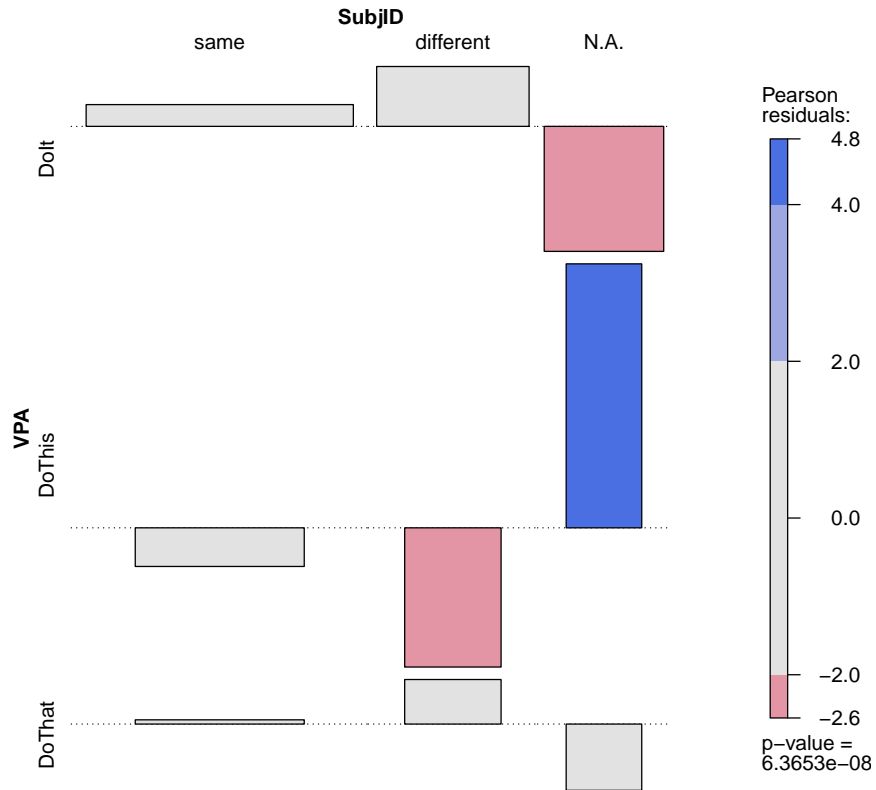


Figure 2.5: Frequency of VPAs having the same or a different subject from the antecedent-trigger

(2013) they do not specifically address the question of identity between subjects, and rely only on invented examples. Both sources argue that *do it* is preferred over *do so* if it denotes the same specific action as the antecedent clause (and not just the same type of action). Quirk et al. provide the pair in (42)–(43) to illustrate this constraint:

(42) Martin is painting is house. He does it every four years.

(43) Martin is painting is house. His neighbour did so last year.

However, corpus data and experimental results from Miller (2013) sug-

gest the opposite constraint, namely that *do so* is preferred if the SoAs in both clauses are identical. Miller also brings evidence that contrasting SoAs can make *do so* strongly dispreferred.

Results

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Same	284 (56.8%)	112 (56%)	117 (58.5%)	513 (57%)
Different	172 (34.4%)	38 (19%)	62 (31%)	272 (30%)
NA	44 (8.8%)	50 (25%)	21 (10.5%)	115 (13%)
TOTAL	500	200	200	900

Table 2.9: Identity of SoAs between the antecedent and VPA clauses

As Table 2.9 shows, the most frequent situation for VPAs is to refer to the same SoA as in the antecedent clause. *Do this* again turns out to be infrequent on contexts where there are different SoAs in each clause, and more frequent than *do it/do that* when SoA identity is undetermined. Thus, although χ^2 -test shows a significant association between VPA choice and SoA identity, this result is a side-effect of the behaviour of *do this*, and does not otherwise show any real influence of SoA identity. Analysis of Pearson residuals confirms this tendency—as with subject-identity before, *do this* is especially frequent in contexts where identity cannot be determined, and especially infrequent with different SoAs across clauses.

2.3.7 Structure of the antecedent-trigger

Previous work

Very little work has been carried out on the structure of VPA antecedent-triggers specifically. This is largely because the literature on ellipsis has long been concerned with the problem of antecedent mismatch with VPE,

2.3. OVERVIEW OF CORPUS DATA

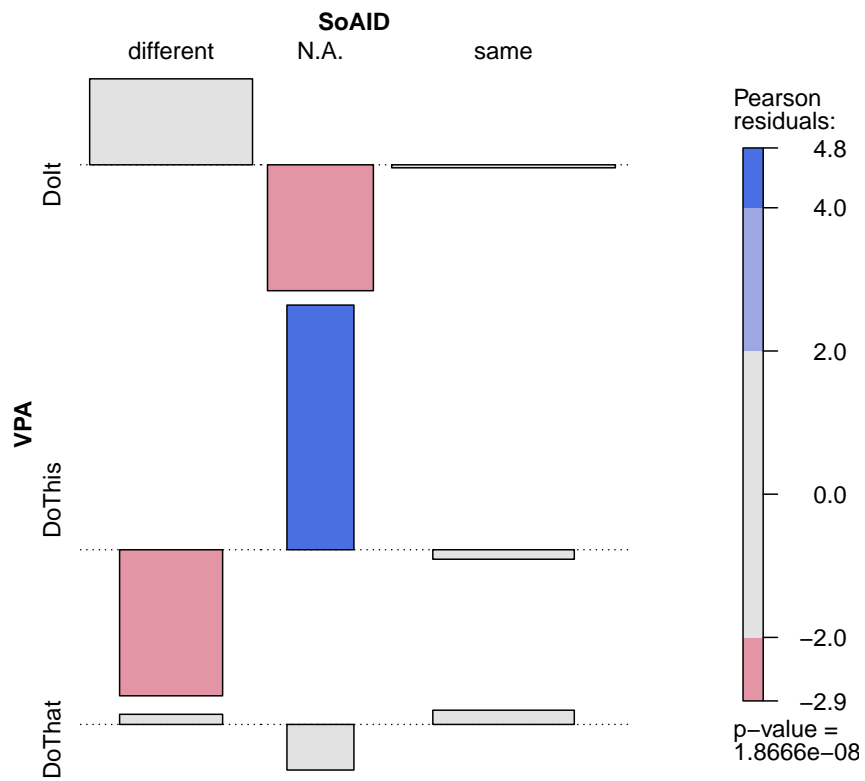


Figure 2.6: Association plot for identity between SoAs in the antecedent and VPA clauses

the acceptability of which has long been a matter of debate. Since studies of mismatch have always assumed that VPAs always allow mismatch, the need has never been felt for detailed analysis of the possible structures through which VPA antecedents could be accessed. Some preliminary discussion of VPAs with nominal antecedent-triggers and inference can nevertheless be found in [Souesme \(1985\)](#) and [Miller \(2011\)](#). Beyond this, there have been no detailed accounts of the types of structures occurring as antecedent-triggers for VPAs, and the role of this feature remains unclear.

Coding scheme

The coding indicates the structure of the antecedent-trigger if it is finite (we code this as `main VP`), or the type of structure under which it is embedded if it is non-finite, such as the complement of a verb, noun, or other. If the antecedent-trigger is something other than an active VP, for instance a passive VP or NP, the type of structure of the trigger itself is indicated rather than the structure containing it (for instance, we do not specify whether the passive VP is a complement or the main VP of a finite clause).

The categories distinguished for the structure of the antecedent-trigger are the same as those used to code the structure of the VPA, with the exception of passive VPs (as we decided to only include active occurrences into the corpus) and impossible categories such as NP or VPE. Therefore, the annotation scheme described below is applicable indifferently to the VPA or antecedent-trigger, except for the categories just mentioned. For the sake of convenience, we will make use of invented examples of VPA for the most simple structures, and attested ones for categories that are only relevant for the antecedent-trigger, or where the structure would prove difficult to illustrate otherwise.

The most simple structure is where the VPA occurs as the main VP predicate of the sentence, e.g., *I did it*, which we code as `main VP`. This

is the default for finite non-anaphoric VPs other than passives, as well as imperatives.

2.3.7.1 Complements

The VP may be the complement of a verb (*I want to do it*), a noun (*a way to do it*), or adjective (*He was very good at doing this*); these structures are abbreviated respectively as *Vcomp*, *Ncomp* and *AdjComp* in the coding.

2.3.7.2 Auxiliaries

We use the label *Aux* when the VP occurs after auxiliary *be*, *do* or *have*, e.g., *I'm doing it*, *Did you do it?*, *I haven't done it*. Modal auxiliaries (*I can do it*) are labelled separately as *modal*. The *Aux* category is in fact part of the broader verb complement class described above, but it is important to distinguish auxiliaries for the purposes of examining alternation with VPE, which can occur after any auxiliary or *to*⁶. The *modal* label, which is also applied to modal verbs such as *have to*, is more a semantic category than a syntactic one, but as will be shown in Chapter 7, the alternation between VPAs and VPE after modal auxiliaries differs in some cases from the behaviour observed with other auxiliaries.

2.3.7.3 Anaphoric constructions

The antecedent-trigger of a VPA may itself be a VPA or another anaphoric construction, namely VPE, Null Complement Anaphora (NCA, see [Shopen 1972](#); [Sag 1976](#)), or a pronoun that is not the complement of *do*. These categories are used only for the antecedent-trigger; for VPAs, the label *VPA* would obviously be uninformative, and all others are impossible.

6. [Miller and Pullum \(2014\)](#) cite [Levine \(2012\)](#) as making the case for treating *to* as an auxiliary.

(44) illustrates an anaphoric chain with the antecedent action referred to first with VPE, then twice with *do it*. The only example of NCA (ellipsis of the entire VP complement) is shown below in (45); (46) shows a case where the antecedent-trigger is a pronoun outside of VPA.

- (44) How could he keep Andrew at Agricultural college? Well they would, somehow or other. They'd done it for Adam and Christopher and they'd do it for him. (AC2)
- (45) He had these weights—dumbbells like—and he said if anybody could lift them over their heads, he'd give them a gold watch. Well, you never saw such a sight. They all tried. Big lads they were and all. Not a one could do it. (ACV)
- (46) Why shouldn't I have a good old weep? I've worked for it, I deserve it and I'm jolly well going to do it. (APU)

2.3.7.4 Non-parallel antecedent-triggers

The antecedent is considered non-parallel if it is anything other than an active VP. It may be a passive VP (47), an NP (48) or it may not even be a constituent (49).

- (47) So, it isn't as if we're saying, erm maths has always been taught quite well but we can do it better. (KRH)
- (48) What are your plans, then? I suppose if you're a writer you can do it anywhere. Must be wonderful. (CKB)
- (49) They're retiring to various parts of the country, down here onto the Sussex coast and on the whole coast into the South West—to the pleasant places if they can afford to do it [...] (KRE)

2.3.7.5 Other non-finite structures

Non-finite cases other than complements are labelled simply as VP[to], VP[ing] or VP[bse] according to the form of the head verb. A more detailed classification distinguishing between various non-complement cases (such as VP[to] subjects or adjuncts, e.g., *It's easy to do it; We hired somebody to do it*) would have been needlessly complex in view of the small number of occurrences in each of the subcategories. Below are some attested examples of such non-finite structures for the VPA (50)–(51) and the antecedent-trigger (52):

- (50) To actually stop on the hard shoulder is very dangerous. But to do it in lane three, to get out and change a tyre is beyond belief. (K6D)
- (51) What I'm trying to say is that it's providing for people already here that I think should be a major item of population policy, and doing it not in a massive world sense, or even a country, but in particular areas as well. (KRE)
- (52) 'So what do we do? Phone up the London police?'
'No use doing that.' (AC4)

2.3.7.6 No antecedent-trigger

The absence of an antecedent-trigger is indicated by NA and is used primarily for exophoric uses.

Results

As Table 2.10 shows, verb complements and finite main VPs are the most common structure for the antecedent-trigger. Notable differences between VPAs are observed with main VPs and cases without an antecedent-trigger, which are much more common with *do this*. *Do it* is by far the most

frequent form to have another VPA as its antecedent-trigger, often another occurrence of *do it*. It is also the only form attested with VPE as the antecedent-trigger.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Vcomp	72 (14.4%)	30 (15%)	35 (17.5%)	137 (15%)
Main VP	55 (11%)	44 (22%)	34 (17%)	133 (15%)
Modal	71 (14.2%)	19 (9.5%)	26 (13%)	116 (13%)
NA	43 (8.6%)	48 (24%)	14 (7%)	105 (12%)
VPA	80 (16%)	5 (2.5%)	8 (4%)	93 (10%)
Aux	34 (6.8%)	16 (8%)	16 (8%)	66 (7%)
VP[to]	29 (5.8%)	13 (6.5%)	14 (7%)	56 (6%)
VP[ing]	25 (5%)	5 (2.5%)	10 (5%)	40 (4%)
Ncomp	23 (4.6%)	6 (3%)	9 (4.5%)	38 (4%)
Passive VP	21 (4.2%)	2 (1%)	7 (3.5%)	30 (3%)
Adjcomp	9 (1.8%)	7 (3.5%)	6 (3%)	22 (2%)
NP	10 (2%)	3 (1.5%)	8 (4%)	21 (2%)
Pronoun	10 (2%)	0	3 (1.5%)	13 (1%)
VPE	12 (2.4%)	0	0	12 (1%)
VP[bse]	4 (0.8%)	1 (0.5%)	7 (3.5%)	12 (1%)
Mixed	1 (0.2%)	1 (0.5%)	2 (1%)	4 (0.4%)
NCA	1 (0.2%)	0	0	1 (0.1%)
AP	0	0	1 (0.5%)	1 (0.1%)
TOTAL	500	200	200	900

Table 2.10: Structure of the antecedent-trigger

The χ^2 -test is not applicable due to the low frequency of a number of structures. Pearson residuals suggest main VPs are more typical of *do this* (2.6) rather than *do it* (-2.2). VPAs are much more frequent than expected as antecedent-triggers of *do it* (3.9) and very rare with *do this* (-3.4) and *do that* (-2.8).

2.3.8 Clause type of the antecedent-trigger

Coding scheme

We code the type of clause containing the antecedent-trigger and VPA, as well as its main or subordinate status. Finite clauses are divided into declaratives (*I want to do it*), direct or indirect interrogatives (*Can you do it?*; *I don't know how to do it*) and imperatives (*Do it!*), whereas non-finite clauses are classified as infinitival (as in (50), for instance) or participial.

(53) Then we had tea together with me sitting up in bed in my dressing-gown. I often did this. (FEE)

Relatives (*The people who do it*), *if*-clauses (*If you do that*) as well as clefts (see (51) above) and pseudo-clefts (54), are labelled as such.

(54) So what you do is, to type check disk space slash and then it asks you wh– if it comes up with a question, which I expect it may well, if you've never done this before. (HDV)

Results

As shown in Table 2.11, declaratives are the most common type of clause for the antecedent-trigger.

The χ^2 -test is not applicable due to the low frequency of certain clause types. However, aside from the already-observed fact that *do this* often occurs without an antecedent-trigger, there are no notable differences between VPAs with respect to the clause type of the antecedent-trigger. Pearson residuals confirm the tendency described for *do this*, and also suggest that it is especially uncommon.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Declarative	266 (53.2%)	96 (48%)	112 (56%)	474 (53%)
NA	44 (8.8%)	52 (26%)	17 (8.5%)	113 (13%)
Infinitival	60 (12%)	13 (6.5%)	22 (11%)	95 (11%)
Participial	43 (8.6%)	11 (5.5%)	10 (5%)	64 (7%)
Relative	27 (5.4%)	13 (6.5%)	9 (4.5%)	49 (5%)
Dir. inter.	22 (4.4%)	4 (2%)	13 (6.5%)	39 (4%)
If-clause	12 (2.4%)	1 (0.5%)	5 (2.5%)	18 (2%)
Ind. inter.	11 (2.2%)	3 (1.5%)	3 (1.5%)	17 (2%)
Imperative	10 (2%)	2 (1%)	2 (1%)	14 (2%)
Pseudo-cleft	1 (0.2%)	4 (2%)	6 (3%)	11 (1%)
Extraposed	1 (0.2%)	1 (0.5%)	1 (0.5%)	3
Cleft	2 (0.4%)	0	0	2
Ind. exclamative	1 (0.2%)	0	0	1
TOTAL	500	200	200	900

Table 2.11: Syntactic type of the antecedent clause (*dir. inter*=direct interrogative, *ind. inter*=indirect interrogative, *ind. exclamative*=indirect exclamative)

2.3.9 Main or subordinate status of the antecedent clause

Coding scheme

The coding indicates whether the antecedent clause is a main or subordinate clause; NA is used if there is no antecedent-trigger.

Results

Table 2.12 shows the distribution of main and subordinate antecedent clauses across VPAs. The antecedent-trigger is most often introduced in a main clause, but this is more common with *do that* than *do it* and *do this*.

The main or subordinate status of the antecedent clause also appears to have no effect on VPA choice: the χ^2 -test gives a significant result only if NA occurrences are included ($p < 0.001$), not otherwise ($p = 0.07$), an effect

2.3. OVERVIEW OF CORPUS DATA

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Main	220 (44%)	79 (39.5%)	107 (53.5%)	406 (45%)
Subordinate	236 (47.2%)	69 (34.5%)	78 (39%)	383 (43%)
NA	44 (8.8%)	52 (26%)	15 (7.5%)	111 (12%)
TOTAL	500	200	200	900

Table 2.12: Main or subordinate status of the antecedent clause

attributable to the frequency of cases where *do this* has no antecedent-trigger. This result suggests that VPA choice is unaffected by whether the antecedent clause is main or subordinate.

2.3.10 Structure of the VPA

Previous work

While the structure of VPA antecedent-triggers has occasionally been discussed in the literature, the role of the structure in which the VPA itself occurs has never been studied, although there is good reason to assume that the immediate syntactic environment of a VPA will have a stronger influence than that of the antecedent-trigger.

Coding scheme

As already mentioned in section 2.3.7, the categories used to code the structure of the VPA are the same as for the antecedent-trigger except for VPE, NCA, passive VP, NP and NA, which are impossible, and VPA, which would be uninformative.

Results

As seen in Table 2.13, VPAs most often occur after a modal, or as the main VP of a finite clause. They are also commonly attested as complements of

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Modal	171 (34%)	28 (14%)	50 (25%)	249 (28%)
Main VP	103 (21%)	43 (22%)	42 (21%)	188 (21%)
Aux	93 (19%)	55 (28%)	30 (15%)	178 (20%)
Vcomp	64 (13%)	38 (19%)	34 (17%)	136 (15%)
VP[to]	22 (4%)	17 (9%)	14 (7%)	53 (6%)
Ncomp	21 (4%)	6 (3%)	10 (5%)	37 (4%)
Adj comp	13 (3%)	9 (5%)	11 (6%)	33 (4%)
VP[ing]	13 (3%)	4 (2%)	8 (4%)	25 (3%)
VP[bse]	0	0	1 (1%)	1 (0.1%)
TOTAL	500	200	200	900

Table 2.13: Structure containing the VPA

auxiliaries or lexical verbs. The main sources of variation between VPAs are modals, which appear to favour *do it*, and other auxiliaries, which appear to favour *do this*. The association plot in Fig. 2.7 confirms this preference, and also shows that the frequency of *do this* after a modal is less than expected.

2.3.11 Syntactic type of the VPA clause

Previous work

Neither the syntactic type of the VPA clause nor its main or subordinate status have been studied in the literature.

Coding scheme

The categories used are the same as for the antecedent-trigger clause except for NA, which is impossible.

2.3. OVERVIEW OF CORPUS DATA

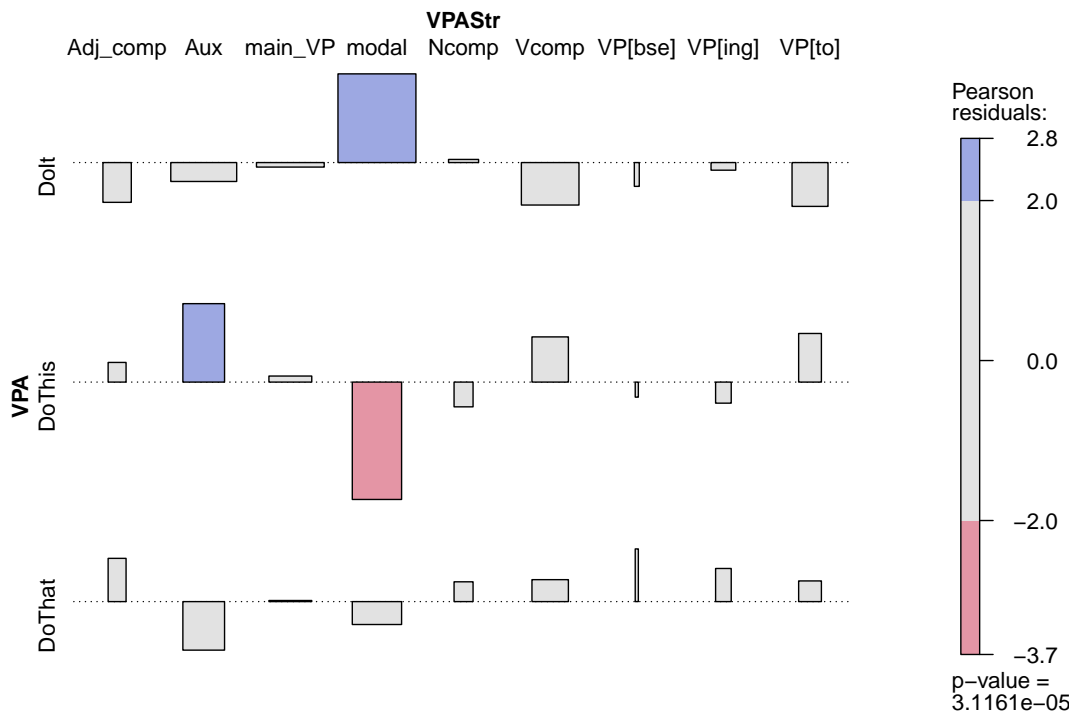


Figure 2.7: Association plot showing Pearson residuals for the structure containing the VPA

Results

As with the antecedent-trigger, declaratives are the most frequent type of clause with VPAs; this is more frequent with *do it* than others. *Do it* is also less frequent in infinitival and participial clauses. Pearson residuals only show higher than expected frequencies for *do that* in infinitival clauses (2), *do this* in participial clauses (2.7), and a lower than expected frequency of *do it* in infinitival clauses.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Declarative	343 (68.6%)	111 (55.5%)	109 (54.5%)	563 (63%)
Dir. inter.	34 (6.8%)	21 (10.5%)	17 (8.5%)	72 (8%)
Infinitival	21 (4.2%)	19 (9.5%)	21 (10.5%)	61 (7%)
Relative	32 (6.4%)	13 (6.5%)	15 (7.5%)	60 (7%)
Participial	17 (3.4%)	20 (10%)	12 (6%)	49 (5%)
Ind. inter.	29 (5.8%)	6 (3%)	4 (2%)	39 (4%)
If-clause	12 (2.4%)	7 (3.5%)	11 (5.5%)	30 (3%)
Imperative	7 (1.4%)	2 (1%)	7 (3.5%)	16 (2%)
Extraposed	3 (0.6%)	0	2 (1%)	5 (1%)
Cleft	1 (0.2%)	1 (0.5%)	2 (1%)	(0.4%)
Pseudo-cleft	1 (0.2%)	0	0	1 (0.1%)
TOTAL	500	200	200	900

Table 2.14: Syntactic type of the VPA clause (*dir. inter*=direct interrogative, *ind. inter*=indirect interrogative)

2.3.12 Main or subordinate status of the VPA clause

Coding scheme

The main or subordinate status of the VPA clause was indicated in the same way as for the antecedent-trigger.

2.3. OVERVIEW OF CORPUS DATA

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Main	267 (53.4%)	86 (43%)	88 (44%)	441 (49%)
Subordinate	233 (46.6%)	114 (57%)	112 (56%)	459 (51%)
TOTAL	500	200	200	900

Table 2.15: Frequency of VPAs in main and subordinate clauses

Results

As seen in Table 2.15, main and subordinate occur about as often overall. However, *do it* is more common in main clauses, while *do this/that* more often occur in subordinate clauses. The association between VPA choice and main/subordinate status of the clause is significant ($p < 0.05$), but Pearson residuals (see Fig. 2.8) show no notable deviation from the expected frequency.

2.3.13 Discourse status of the antecedent

Previous work

Quirk et al. (1985, p. 877) suggest that the information status of the antecedent plays at least some role in VPA choice. They explain that the pronoun in *do that* can receive prosodic stress (the same goes for *do this*, although Quirk et al. do not discuss it in this respect), thus signalling information that is ‘treated to some extent as new or contrastive’, whereas *do it* conveys ‘entirely given’ information.

Coding scheme

The antecedent is considered as discourse-new if it has not been mentioned before the antecedent clause, and discourse-old if it has. We code this as `new` and `old` accordingly. In exophoric cases, the antecedent is discourse-new since the VPA sentence is the first mention of the action in discourse;

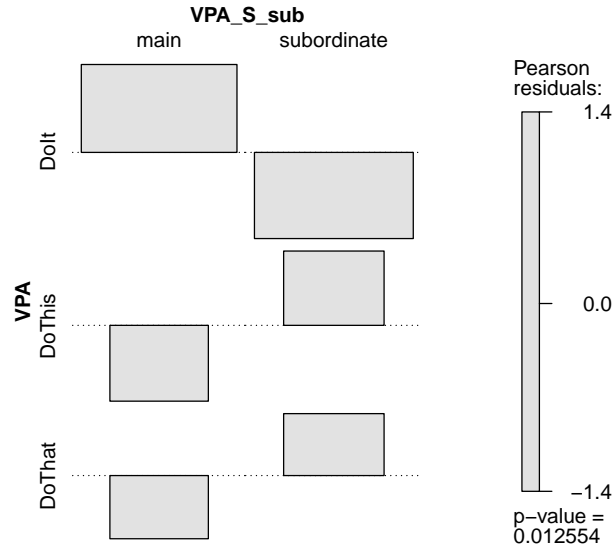


Figure 2.8: Association plot for the frequency of VPAs in main and subordinate clauses

antecedents retrieved from contextual inference are treated as discourse new. This means that uncertain cases (NA) are far less frequent.

Results

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Old	292 (59.8%)	31 (16.2%)	53 (26.5%)	376 (43%)
New	196 (40.2%)	160 (83.8%)	147 (73.5%)	503 (57%)
TOTAL	488	191	200	879

Table 2.16: Discourse status of antecedent

The association between VPA choice and discourse status of the antecedent is highly significant ($p < 0.001$). Residuals show that *do it* is especially frequent with discourse-old antecedents and infrequent with

discourse-new antecedents. *Do this* and *do that* show the opposite pattern, but the effect is weaker for *do that*. This seems in keeping with the contrast between *it* and demonstratives with respect to the information status of the referent.

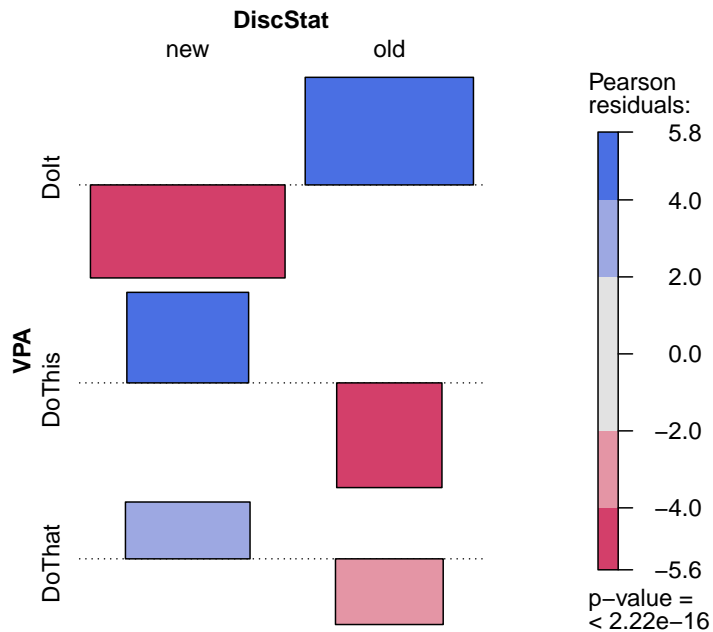


Figure 2.9: Association plot for the distribution of VPAs with discourse-old/new antecedents

2.3.14 Agentivity

Previous work

Agentivity has often been cited as a semantic restriction on the antecedent of VPAs, including *do so* for some authors. The most radical statement of this constraint is made by [Culicover and Jackendoff \(2005\)](#) who consider non-agentive uses (including stative antecedents) to be only possible with

VPE. [Huddleston and Pullum \(2002\)](#) consider them as acceptable with *do so* but not *do it/this/that*; [Souesme \(1985\)](#) and [Quirk et al. \(1985\)](#) make this claim for *do it* only and do not say whether *do this/that* must also be agentive, but they qualify their claims by adding that the restriction holds only ‘for some speakers’ (p. 877). However, at least one non-agentive case is reported by [Miller \(2011\)](#):

- (55) Although straight narrative can advance a plot faster and less circuitously, dialogue often does it more interestingly. ([Miller, 2011](#), ex. 1b, p. 83)

Coding scheme

The term ‘agentive’ is used here (and by the authors cited) in the sense of [Gruber \(1967\)](#), i.e, volitional: the action is wilfully performed by an animate agent.

Results

Preliminary data in [Table 2.17](#) confirm the preference for agentive antecedents, but also reveal that non-agentive ones are at least a marginal possibility. *Do this/that* appear to allow non-agentive antecedents more often than *do it*, but the significance of this effect is difficult to assess given very the small number of non-agentive tokens.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Agentive	496 (99.2%)	191 (95.5%)	193 (96.5%)	880 (97.8%)
Non-agent.	4 (0.8%)	9 (4.5%)	7 (3.5%)	20 (2.2%)
TOTAL	500	200	200	900

Table 2.17: Frequency of agentive and non-agentive VPAs

Pearson residuals (see [Fig. 2.10](#)) suggest *do it* is dispreferred if the antecedent is non-agentive, whereas *do this* allows non-agentive uses more

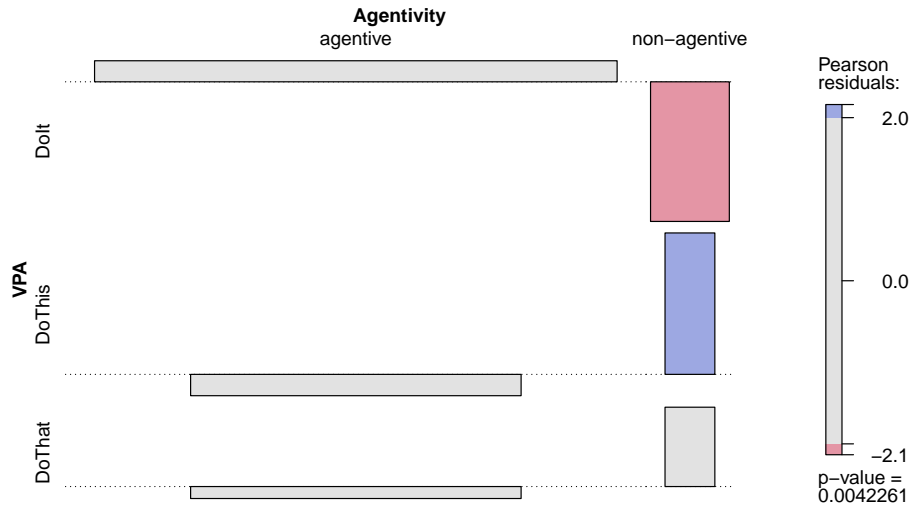


Figure 2.10: Association plot for the distribution of VPAs with agentive and non-agentive antecedents

easily. *Do that* appears to be less indifferent to the agentivity of the antecedent.

No serious conclusions can be drawn regarding agentivity at this stage, but the observed effect deserves to be investigated further. This will be the focus of Chapter 4, which will reconsider the role of agentivity on the basis of a larger sample of non-agentive cases, collected using special heuristics in BNC queries.

2.4 Conclusion

The survey of the corpus data conducted here has allowed us to identify the following features as potentially relevant parameters of VPA choice: register, the presence or absence of an adjunct, the discourse status of the antecedent, and, to a lesser extent, the structure containing the antecedent-trigger and VPA. Agentivity also appears to play a role in VPA choice, but

its influence remains largely unknown at present.

Do it is more common in spoken discourse, prefers to occur followed by an adjunct, and normally refers to actions that have already been mentioned before the antecedent clause, as evidenced by the frequency of VPA and VPE antecedent-triggers. *Do this* mainly occurs in writing, and typically refers to actions that have not been mentioned in the antecedent clause. It is also the most frequent form to occur without an antecedent-trigger. *Do that* strongly prefers to occur without a subsequent adjunct, and shows a similar but weaker preference than *do this* for discourse-new antecedents.

Chapter 3

Compositionality

3.1 The compositionality problem

The present chapter argues for a compositional analysis of the VP anaphor *do it* as having the structure *do* + NP, more specifically main verb *do* + pro-VP *it*. Following [Simner \(2001\)](#) and [Stroik \(2001\)](#), it is argued that *it*, rather than *do it* as a whole, constitutes the anaphoric element. Drawing on [Cornish \(1999\)](#)'s distinction between antecedent and antecedent-trigger, I review both syntactic and semantic evidence supporting the idea that *do it* is compositional. I then discuss how the analysis can be extended to *do this/that* in order to account for the alternation between these forms and *do it*.

3.1.1 The internal structure of VP anaphors

The internal structure of *do* + pronoun VP anaphors—*do this/do that*—has rarely been studied in the literature, in contrast to the numerous and diverse proposals made on the structure of VPE. At least a few studies have addressed the structure of the closely-related form *do so*, which is much less transparent ([Bolinger, 1970](#); [Lakoff and Ross, 1976](#); [Higgins,](#)

1992; Cornish, 1996; Houser, 2010; Miller, 2013), but the set formed by *do it/do this/do that* has never been discussed in detail. All of these forms are usually regarded as mostly similar from a syntactic point of view, and semantically interchangeable, so that a detailed account of their structure or discourse properties has never been deemed necessary.

Most of the work on *do it* has actually been concerned with assessing the claim put forward by Hankamer and Sag (1976) that there are two classes of anaphors: surface anaphors, which require a linguistic antecedent, and deep anaphors, which do not. This hypothesis initially stemmed from the observation that some anaphors, among which *do it*, can have syntactically non-parallel or ‘mismatched’ antecedents as well as exophoric uses, while others, such as VPE, apparently do not allow this. Based on these assumptions, *do it*, a deep anaphor, has commonly been compared to VPE, a surface form, to examine their behaviour with regard to purported deep/surface anaphor features. The ultimate aim, however, was mainly to analyse the structure of VPE, in particular to determine whether there is any structure present at the ellipsis site. Interest in this question was accrued by the discovery of attested examples of VPE violating the above-cited constraints on surface anaphora (see e.g., Hardt, 1993), and therefore questioning the surface anaphor status of VPE. Psycholinguistic studies on this issues have obtained mixed results, either supporting the deep surface status of VPE (Tanenhaus and Carlson, 1990) or finding no evidence for it (Murphy, 1985), and in some cases even questioning the validity of Hankamer and Sag’s distinction. *Do it*, on the other hand, is generally left out of consideration, and no attempt is made to describe its structure beyond the fact that it has the usual properties of a deep anaphor. It is only rarely mentioned outside of the debate surrounding the deep/surface distinction, and has never been studied independently of problems related to the structure of VPE. This chapter proposes to bridge the gap by examining the structural properties of *do it* in relation to its

discourse usage. The next section reviews a small set of mostly liminal proposals that have been made concerning the structure of *do it* and discusses their implications for the compositional analysis defended here.

3.1.2 Literature review

Although some studies of VP anaphors comment in passing on the structure of *do it*, there are very few that directly make a case for or against compositionality. However, those studies which address the question of compositionality all argue in favour of such an analysis; there are none, to my knowledge, that have defended exactly the opposite. Proponents of a compositional analysis include [Simner \(2001\)](#); [Stroik \(2001\)](#) for *do it*, and [Miller \(2013\)](#) for *do it/this/that* ([Miller](#) argues that compositionality is what distinguishes *do it/this/that* from *do so*).

[Simner \(2001, p. 58\)](#) is one of the most explicit proponents of a compositional approach. Criticising the ‘traditional’ analysis, as she calls it, which views *do it* as a pro-VP, she instead advocates a structure in which the pronoun alone is anaphoric and *do* is main verb. [Simner](#) insists (p.58) that in this latter structure “although the ‘do it’ expression is syntactically a VP, it is anaphoric only in part (i.e. only the pronoun is an anaphor)”. In support of this claim, she cites independent evidence that *it* can derive its meaning from a previous VP without the verb + *it* sequence being treated as anaphoric. For instance, the second clause of (56) would not be regarded as involving ‘*try it* anaphora’:

- (56) Mary dived into the pool. John wanted to try it too. (ex. 17, p. 63)

Likewise, [Stroik \(2001, p. 364\)](#) openly rejects a non-compositional view of the structure and insists that in *do it* and *do so*, “it is the object proform *it/so*, not the *do so/it* constituent, that replaces the main VP”. In a discussion of the light verb status of *do*, he argues in essence that *do it* is

formed of helping verb *do* and a VP proform, stressing that it is only the pronoun *it*, rather than *do it*, which replaces the VP in this construction. In support of this analysis, [Stroik](#) cites evidence from answers to *wh*-questions with *do* ('What did X do?'), and the possibility for the VP to serve as the antecedent of a relative, or undergo pseudo-clefting ([Ross, 1970](#)):

- (57) A—What are you doing?
B—Eating the leftovers.
- (58) Ted left, which he shouldn't have done.
- (59) Pat has read a book, which is what Sam had done too.

These data are also mentioned by [Quirk et al. \(1985\)](#), who consider *do it* and *do that* as 'straightforward verb + direct object constructions', in other words similar to the pattern found with any other transitive verb. In this respect, they also note that *do it/do that* can be passivised, as shown in (60) for *do it*:¹

- (60) If, if I walked up to a policeman in the street and gave him a little shove, the chances are he would arrest me, unless it was done in a totally friendly way. (KRH)

In most other studies, even though the question of whether or not VPAs are compositional is left out of consideration, it is still possible to determine whether an author is assuming that *do it/this/that* have a compositional structure. As we will see, the analysis defended here is in fact

1. Quirk et al's original example with *do that* is as follows:

- (1) A: Have you noticed the front wheel is buckled?
B: That was done ages ago

This is to some degree unnatural since the antecedent is stative and does not imply a previous action performed on the object (e.g., twisting the wheel or other). For this reason, *do that* as well as *do it/do this* are slightly awkward here and a more natural answer would be along the lines of *That/It happened ages ago*, consistently with the non-agentive interpretation.

3.1. THE COMPOSITIONALITY PROBLEM

the dominant one (insofar as *do* and the object pronoun are regarded as separate entities), even if it is most often not openly stated.

Some authors, like [Hankamer and Sag \(1976\)](#); [Culicover and Jackendoff \(2005\)](#), appear to adopt a non-compositional view, considering that *do it* as a whole functions as a pro-VP, similarly to *do so* and *do this/do that*. Such descriptions make no reference to a possible antecedent for *it*, either syntactically (an actual linguistic structure) or semantically (a possible interpretation). In the absence of any discussion of compositionality, it is hard to decide whether the authors cited would actually reject this hypothesis, but overall their analyses seem to follow from the opposite assumption. [Culicover and Jackendoff \(2005, p. 126\)](#), for instance, write that *do it* and other *do* + proform anaphors (collectively referred to as ‘do X anaphora’) are pro-VPs that stand in for ‘a subpart of a VP—including the verb itself and possibly non-contiguous portions of the VP’, allowing for remnant complements or adjuncts, e.g., *Robin broke the window (with a hammer) and Mary did it to the table top (=broke the table top (with a hammer))*. Although they comment on the ‘nontrivial semantic content’ of the proform in expressions such as *do something/do the same/do the opposite*, they do not specify whether this also holds for *do it/this/that*.

Other analyses appear to argue (implicitly at least) in favour of a compositional structure. This appears very clearly in the work of [Cornish \(1992\)](#), who discusses *do so/do it* in a more general study of the anaphoric properties of *so* and *it* across a range of host verbs, (also including e.g., *say: say so/say it*). The discussion in [Cornish \(1999, p. 84\)](#), although not in any way centred on *do it*, still comes close to an actual analysis of the structure, proposing that *it* in *do it* refers to an action and presenting various paraphrases of this sense depending on the context. In the following, for instance, Cornish proposes that the antecedent can be expressed as ‘a specific, definite act of shooting down at least one ICBM in space’:

(61) The Americans have developed a new method of shooting down

intercontinental ballistic missiles [ICBMs] in space : they did it on Sunday.

(BBC Radio 4, 12 June 1984)

The antecedent of *it* varies with features such as tense or aspect, so that the progressive *they were doing it...* yields a reading where *it* refers to “an (unspecified) number of acts of shooting down ICBMs (more than just a single one) in space”. Although Cornish does not explicitly describe *do it* as being compositional, this is certainly the approach he is following here.

Similarly, Schuster (1988) discusses *do it/this/that* in the context of a study on the use of *it* and *this/that* with action referents, of which VPAs are taken to be just one instance. This is apparent in her notation of anaphoric relations, reproduced in (62) (the exchange occurs between an Emacs instructor and a novice user performing basic tasks with the software; interactions are carried out via a computer terminal):

- (62) U: Tell me how to [define a region]_i again. I forgot how to do *it*_i.
E: [You set the mark with <esc>-M at one extreme of the text and you move the cursor to the other extreme]_j. *That*_j defines a region.
(ex. 2, p. 1)

Here, as in Cornish’s example above, *it* in *do it* is clearly assumed to be the only anaphoric element, although again no particular justification is provided for this description. Also, neither author explicitly claims that this is the correct analysis of VPAs (the term is never used at all in their accounts).

A similar view is held by Souesme (1985), but the status of *it* is rather more loosely defined as referring simply to ‘something mentioned before’ (p. 35). Merchant (2004) seems to make similar assumptions on the structure, but does not conclude on the status of *it* (‘whatever *it* refers to here’, p. 720).

Simner et al. (2003) point out that examples like (63), where the antecedent of *do it* is of the form verb + NP object and the NP can serve as

3.1. THE COMPOSITIONALITY PROBLEM

a complement of *do*, are ambiguous between a ‘nominal’ reading where *it* refers to the NP object, and a ‘verbal’ one where *do it* as a whole is interpreted in terms of the previous VP, in this case ‘borrow the puzzle’:

- (63) Sam borrowed a jigsaw puzzle, and he *did it* when everyone was out.

[Simner et al.](#) are chiefly concerned with the contextual factors influencing the choice of one or the other interpretation. Using experimental evidence, they investigate how *do it* is interpreted as ‘verbal’ *do it* or as *do* + pro-NP *it* depending on whether a sentence boundary or turn intervenes between the antecedent and anaphor, or if a subject pronoun follows the conjunction in coordinated sentences. However, their description of the structure is minimal. It is simply argued that VPA *do it* involves *do* and a ‘deverbalized NP’, in this instance [_{NP} (*that thing that is*) *the borrowing of a puzzle*], as opposed to a ‘simple NP’ when *do it* occurs in the so-called ‘nominal’ use, where *it* takes an NP antecedent.

The various analyses discussed above converge in treating the *it* of *do it* as an independent proform which takes the VP as its antecedent, hence a pro-VP. Another recurring argument is the non-idiomatic nature of *do it*, and the idea that it is not a special use of *do* but an ordinary transitive pattern. [Huddleston and Pullum \(2002, p. 1532\)](#) similarly argue that *do it/this/that* are not idioms (contrary to *do so*, which, as they explain, is not directly analysable as *do* + *so*). It follows that *do it*, in their view, involve the usual main verb, which they call ‘general agentive’ (do_{ga}), the same as in any other *do* + NP construction, e.g., *I’ve just done something very stupid*. The analysis of *it* is less detailed, however. The data seen above for the passive and relative clauses are mentioned briefly, but [Huddleston and Pullum \(2002\)](#) do not elaborate on what the antecedent of *it* might be. Presumably *it* refers to an action, as proposed for *do this/do that*, for instance, *He was doing this* could be uttered ‘as I demonstrate the action in question’.

In summary, the consensus seems to be that *do it* is made up of main verb *do* and pro-VP *it*, implying that its properties can be predicted directly from those of *do* and *it*, as indeed [Huddleston and Pullum](#) acknowledge. Since the properties of *do it* are not well-known, however, this issue is not considered most of the time. The next section discusses how the proposed analysis can account for these properties, and presents further evidence for the compositional nature of *do it* as well as *do this/do that*.

Before reviewing these data, a word on must be said on *do so*, which has been left out of the analysis so far. Although it was suggested earlier that *do so* differs from *do it* in being non-compositional, several accounts have argued otherwise. Stroik, for instance, openly rejects the idea that *do so* is a VP proform, arguing instead that it shares the *do* + pro-VP structure of *do it*. While such an analysis is defensible, it seems less convincing to treat *so* as the object of *do*, as Stroik proposes, since *so* is not an NP unlike *it*. [Bouton \(1970\)](#)'s suggestion that *so* is a manner adverb is also unconvincing, since there does not seem to be any manner interpretation associated with *do so*. Note in particular that, as mentioned in section [4.5.7.2](#) above, *do it* often serves as the support for an additional manner comment provided by an adjunct, see table [\(2.7\)](#).

As for the relation between *do so* and *do it*, there are a number of important differences between them. First, while the syntactic category of *so* is still a matter of debate, *it* in *do it* is very clearly an NP. Second, *do so* allows non-agentive and under certain conditions stative antecedents (especially when it is nonfinite and VPE is not available, see [Houser \(2010\)](#); [Miller \(2013\)](#) amongst others for examples and discussion). By contrast, non-agentive antecedents of *do it* are rare and stative ones have not been encountered so far. All this shows that the semantic influence of main verb *do* is not present in *do so*, or in any case not as perceivable as in *do it*. For these reasons it is doubtful that *do so* is compositional in the same way as *do it*, as [Miller](#) argues.

Another point is that the pro-VP status of *so* in *do so* must be distinguished from other anaphoric uses discussed by Cornish (1992) (e.g., *I think/doubt so, he said so*, etc.) where it clearly stands in for a sentential complement (e.g., *I doubt he will come/He said he would come*). Comparison with these other cases may be helpful but is probably not sufficient to account for the properties of *do so*.

3.2 Evidence for compositionality

This section discusses evidence that *do it* is a compositional VP anaphor having the structure *do* + NP. Section (3.2.1) investigates the difference between the two interpretations studied by Simner et al. (2003), proposing that both have the same structure and are distinguished by the pro-VP or pro-NP function of *it*. The existence of these two readings, as well as the alternation between *it/this/that* in VPA and the fact that all of these pronouns can serve independently as pro-VPs, strongly suggest that *do it* as well as *do this/that* are compositional in structure. The syntactic tests mentioned by Stroik and others also support this view. Section (3.2.1.6) focuses more closely on the role of *do* and argues it can be analysed as a main verb in the VPA use just as in the ‘nominal’ (NP anaphora) reading. This status is the same as in any other *do* + object configuration (other than idioms), and is evidenced in particular by the semantic constraints placed by *do it* on its antecedent—non-stative and typically agentive—which correspond to the semantics of main verb *do*.

3.2.1 *It* as pro-VP

3.2.1.1 Ambiguity revisited

The previous section reviewed a range of distributional facts which suggest that *it* is a pro-VP when used in conjunction with *do* for VP anaphora. Here

CHAPTER 3. COMPOSITIONALITY

we review these data in greater detail and show how they support the pro-VP account. Let us first reconsider the two interpretations distinguished by Simner et al for *do it*, given in (64a) -(64b) below:

- (64) a. VPA: Sam borrowed a jigsaw puzzle, and he did it while the others were out.
b. NPA: Sam borrowed a jigsaw puzzle, and he did it while the others were out.

As discussed earlier, Simner et al explain that in the VPA reading *do it* takes its interpretation from the entire VP *borrowed the puzzle*, while in the NPA case it is understood as *do* + anaphoric object, in this instance *did the puzzle*. Of course, not all instances of *do it* are ambiguous in this way: (65) is clearly a case of VPA, since *did it* cannot be taken to mean **did hot afternoons*²:

- (65) Bonnard painted hot afternoons in mirrors. Magritte did it by playing with paradox: day and night both; picture and landscape both.
(BNC:A08)

The two readings are available only when the previous sentence contains both a suitable NP antecedent for *it* and a VP that could be the antecedent of *do it*. This is the case in (64), but not (65), which is therefore unambiguous. Since the nominal reading boils down to the pattern *do* + object, it seems arguable at first sight that the verbal reading is, by contrast, not compositional. Yet as noted, Simner et al. ultimately ascribe the ambiguity of sentences like (64) to the pronoun *it*, which refers either to a simple NP in the *do* + object reading or a ‘deverbalized’ NP in the VPA reading. Their paraphrase of the referent in this interpretation (*that thing*

2. Simner et al. give the following example as a case of unambiguous VPA:

- (1) John swam the English Channel and he did it in under 12 hours.

This is not entirely convincing however, as it is also possible to read *did it* here as *did the English Channel*, with a roughly equivalent meaning.

that is) the borrowing of a puzzle) is not very natural, but it gives an idea of the way *do it* is resolved if *it* is assumed to be the anaphoric complement of *do* in the VPA reading.

A potential problem for the analysis at this stage is that if *it* functions a pro-VP, then it should be possible to restore the VP antecedent as the complement of *do*. This proves impossible with any non-finite form of the VP, either the gerund, the base form or *to*-infinitive, as shown in (66) (assuming *do* is main verb and not the auxiliary):

(66) *Sam did borrowing/(to) borrow the puzzle.

The ungrammaticality of **do* + VP is hard to explain, in particular for the gerund, which is an NP in external structure (cf. Pullum 1991, hence of the appropriate category for a direct object. Since there is generally no problem with replacing *do it* by its VP antecedent, a non-compositional account may seem justified in the light of these facts. However, such an explanation would not easily account for the passive or pseudo-clefting data we have discussed, or the alternation between *do it* and *do this/that*. I return to the case of (66) in Section 3.3, where it is argued that although surprising, these data do not run counter to the analysis defended here takes, which proposes that both readings identified by Simner et al. correspond to the same structure *do* + NP, and that it is really the pronoun *it* which is ambiguous between a pro-VP and pro-NP reading in sentences like those in (64). The VPA use of *do it* is compositional in the same way as the *do* + pro-NP *it* use, and is possible provided that a suitable antecedent is present in or inferable from the context.

3.2.1.2 Properties of the VPA/NPA readings

Simner et al. provide no reliable way of determining if a particular instance of *do it* is VPA or NPA, in their terminology, other than the two types of NP referents for *it* mentioned above. The point of their study is

to show how discourse factors, such as causal relationships or parallelism between antecedent and anaphoric clause, can influence the choice of one of the two readings in ambiguous cases. In practice however, truly ambiguous examples are rare, as it is uncommon to have both an appropriate nominal and verbal antecedent for *it* in context. Below we discuss two main features of VPA *do it* that distinguish it from the *do* + nominal *it* use.

Do So Whereas *do it* is potentially ambiguous because of the two possible types of antecedent for *it*, *do so* is unambiguously a VPA, since *so* cannot be a pro-NP. As such, it only alternates with the VPA use of *do it*, so that the only way to interpret *do so* when it is substituted for *do it* in Simner et al's example sentence is as *borrow the puzzle* (and not *do the puzzle*):

(67) Sam borrowed the puzzle, and he did so when the others were out.

Alternation with *do so* therefore provides one way of telling apart the VPA and NPA uses: if it is possible, then *do it* is VPA under that interpretation, if not, then *do it* is NPA.

Lack of an NP antecedent for *it* It was argued above that the NPA reading is available only if a suitable NP antecedent for *it* can be recovered from the context and can function as the object of *do* (such as *the puzzle* in 64b). In attested examples of VPA *do it*, this is frequently not the case, and a pro-NP reading of *it* is logically ruled out, as in (68) where the antecedent is headed by an intransitive verb:

(68) Big boys don't cry. If you want to do it, do it in the toilet where no one can see you (COCA).

A similar situation arises when the VP antecedent is of the form *do* + plural NP, as in (69) below. In this case *it* cannot refer to *the dishes*, and the VPA reading is the default:

- (69) It's John's turn to do the dishes, but he really doesn't want to do it.

This shows that if no NP suitable antecedent can be recovered from context, *it* must be interpreted instead in terms of the VP. The next section brings further justification for this analysis, showing in particular that the pro-VP use of *it* exists independently of *do it* anaphora.

3.2.1.3 Evidence for the pro-VP status of *it*

The pro-VP function of *it* is by no means restricted to VPA *do it*. In fact, as Cornish points out, *it* can have various types of antecedent alongside NPs, for instance a VP or an S:

- (70) Sam borrowed the puzzle. It [= borrowing the puzzle] was his idea.
- (71) Sam borrowed the puzzle. It [= that he borrowed...] annoyed the others.

Cornish also gives the following examples where *it* stands in for an AP/PP predicate:

- (72) John isn't usually worried, but today he certainly looks it. (11b, p. 167)
- (73) A:— You look on top of the world today!
B:— Kind of you to say so; I really feel it too! (12b, p. 167)

Therefore the pro-VP use is just one of several possible functions for *it*. That this is the function involved in *do it* is shown by a range of data which are indicative of an anaphoric relationship between *it* and the VP.

First, passivisation is possible for VPA *do it* (75), and in that case the antecedent of the subject pronoun is clearly the VP, as indicated in the brackets. Note that in this configuration there is no problem replacing the pronoun by its antecedent. This is exactly similar to *do* + nominal *it* (74):

CHAPTER 3. COMPOSITIONALITY

- (74) The work is extremely urgent. It [the work] has to be done by tomorrow.
- (75) Edmund Hillary was the first man to climb Mount Everest. It [climbing Mount Everest] had never been done before.

These data illustrate the structural parallelism between the two uses of *do it*: in each case the pronoun corresponds to the object of *do*, and is interpreted by retrieving an appropriate antecedent, either an NP or VP. Both types of antecedent can be restored in subject position of a passive use of *do*. These facts are not expected if *do it* is taken to be non-compositional.

Further evidence that *it* is pro-VP comes from the dual behaviour of pseudo-clefts involving main verb *do*, e.g., *What X did was...*, which can target either a VP or an NP:

- (76) a. What Sam did was [_{NP} the puzzle]
b. What Sam did was [_{VP} borrow the puzzle]

In (76b), the VP *borrow the puzzle* is identified with *what*, which corresponds to *it* (and therefore to the direct object of *do*), in the same way as the NP *the puzzle* in (76a). This identificational function of the pseudo-cleft structure again confirms that the VP is the antecedent of *it*. Sentences of this kind are used by Halliday (1967) to assess the relationship of an action with the sense of main verb *do*, and by Cruse (1973) to determine whether a VP is agentive (on the assumption that *do* is itself an agentive verb). Essentially, if the resulting pseudo-cleft is odd, the referent of the VP is not a good example of ‘doing something’ (not a typical ‘action’). Non-agentive processes (77a) and states (77b) are examples of this; they show reduced acceptability in *do*-pseudoclefts, and can be considered as weakly related (or completely unrelated, in the case of states) to the sense of *do*:

- (77) a. #? What John did was get sick.
b. # What John did was look healthy.

3.2. EVIDENCE FOR COMPOSITIONALITY

This is consistent with the fact that non-agentive processes and states make bad antecedents for *do it*—or more rightly for *it* used as the complement of *do*, as noted earlier.

Finally, it is also possible for the VP to be the antecedent of a relative pronoun (Ross, 1970) or the answer to an interrogative pronoun (Cruse, 1973; Souesme, 1985; Stroik, 2001). The interrogative data show that VPs as well as NPs constitute acceptable (and informative) answers to questions of the type *What is X doing?*:

- (78) Ted left, which he shouldn't have done. (cited in Stroik, 2001, p. 364)
- (79) What are you doing?
a. Eating the leftovers.
b. The dishes. (Stroik, 2001, p. 364)

Stroik takes this as evidence that *it* is the element replacing the VP in *do it*, but the identification between the pronoun and the VP is less direct than in the previous tests since the answer need not consist only of the VP or NP. Like pseudo-clefts, interrogatives also reflect the semantic preferences of *do it*, with inappropriate antecedents (e.g., a state in (80)) being less natural answers:

- (80) A. What will he do next ?
B. *He'll like chocolate (Souesme, 1985, ex. 24l, his judgement)

The tests reviewed so far bring evidence of an anaphoric relationship between *it* and the VP antecedent, confirming the pro-VP status. The next section shows that the description proposed here is also valid for the pair *do this/do that*, which share the same structure *do NP* and are semantically close to *do it*, but have different discourse properties.

3.2.1.4 Choice of pro-form: *do it* and *do this/that*

Alternation between *do it* and *do this/that* is a strong argument in favour of the compositional analysis, as it suggests the pronoun is syntactically independent from the entire VP, so that the choice is really between *it/his/that*, rather than anaphoric *do* + pronoun constructions. Crucially, the possibility of replacing *it* by *this/that* is not expected if *do it* is considered non-compositional (note in this respect that *do this/that* lack the idiomatic uses of *do it*, e.g., the ‘success’ sense: *That did it!* vs *#That did this/that!*). This section shows that *do this/that* lend themselves to the same analysis as the one which has been defended so far for *do it*.

First, it can be seen that *this/that* retain their usual distribution when used in *do this/that*. They can be stressed (81a) or modified by adverbs like *just, precisely* etc. (81b) occurring between *do* and the pronoun. These properties are the same as when *this/that* occur with any other transitive verb, e.g., *say*. None of this is possible with *it*, whether the preceding verb is *do* or another transitive verb:

- (81) a. You shouldn’t do (say) THIS/THAT/*IT!
b. I’ll do (say) just this/that/#it.

Both *do this* and *do that* can be passivised, with the subject pronoun taking the VP as its antecedent. In inverted pseudo-clefts of the type ‘This/that is what X did’, the VP antecedent of *this/that* is identified with *what*:

- (82) a. In practice it is necessary to classify these soils. This is done by relating the soil to the type of cleaning agent...(BNC:APV)
b. I shall not go into details about his speeches; that was done at great length on 12 June (HHW)
- (83) a. In other words I was free to speculate and this is what I did. (ASN)

- b. So what do I do? I use my brains, laddie, that's what I do.
(CH4)

These facts are akin to what was observed earlier with *do it*, and support the idea that it is identical in structure to *do this/that*.

3.2.1.5 *Do it and do this/that*: discourse properties

Another aspect of the data concerns the discourse use of *do it* vs *do this/that*. Very little has been said regarding this issue in the literature, on the assumption that all three forms were more or less interchangeable and semantically equivalent.

Although the properties of *do it/this/that* are still unclear at this stage, there is evidence that they are not equivalent in discourse. Preliminary discussion of this issue in Miller (2011) suggests that the alternation between the three forms is driven by more general constraints on the use of demonstrative pronouns and *it*, in particular those proposed by Gundel (1993) who claim that *it* requires a more 'given' antecedent (i.e, more salient to the addressee) than *this* or *that*. Specifically, in their terminology, *it* requires an 'in focus' antecedent (presumed to be in the addressee's short-term memory and at the current centre of his/her attention) while pronominal *this/that* require only an 'activated' one (present in short-term memory, but not in focus of attention). These requirements are exemplified in (84a)–(84b) since the first occurrence of *do it* or *do this* is enough to make the antecedent salient, subsequent reference to the antecedent with a demonstrative (with *this* as in the example, or *that*) is unnecessary. It is therefore infelicitous to repeat *do this* (84a) or reverse the order of *do this* and *do it* in (84b).

- (84) a. I didn't know I couldn't do this ...so I just did it/#this! (ex. 11d, p. 90)
b. I didn't know I couldn't do it ...so I just did #this!

These data highlight differences in the usage of *do it* and *do this/that*, alongside the distributional differences described earlier, and lend further support to the idea that the choice between them is ultimately reducible to the choice between *it* and *this/that*.

3.2.1.6 Main verb status of *do*

Both [Quirk et al. \(1985\)](#) and [Huddleston and Pullum \(2002\)](#) argue that *do it* involves main verb *do*. [Miller \(2011\)](#) also notes that this is an important difference between *do* + proform VPAs and ellipsis with *do* which involves the auxiliary. This is shown by the absence of (subject-verb) inversion with *do it* (**Did he it?*) and the necessity of *do*-support for negation (*Don't do it!*). The difference between main verb and auxiliary is also reflected in the semantic properties of VPE, which easily allows stative and non-agentive antecedents, in contrast to *do it*:

- (85) a. John looks healthy. I wonder why he does#does it.
b. # John got sick. I wonder why he did/#did it

Do it also shows a preference for animate subjects, which are more prototypical agents than inanimates (strictly speaking, it is common to have so-called 'pseudo-animates', i.e. inanimates whose properties bring them close to real animates, for instance machines or 'natural agents' like the wind, sun, rain or other, see Chapter 4). In the following example, based on [Cruse \(1973\)](#), having an inanimate subject in addition to the non-agentive antecedent (unaccusative *break*) adds to the badness of the sentence:

- (86) # The vase broke. It did it after falling from the table.

The same judgement applies to Cruse's original pseudo-cleft sentence:

- (87) # What the vase did was break.

3.3. APPARENT PROBLEMS FOR THE COMPOSITIONAL ANALYSIS

Although attested examples of *do it* with non-agentive antecedents can be found, they are rare (2 in a corpus of 500 occurrences from the BNC) and most occurrences conform to the pattern where the subject is an animate agent. Exceptions involve cases where the subject is at least a cause or inanimate controller of the action, but none where the subject is a patient instead of an agent. The restriction against stative antecedents seems much stronger, as no examples have been found in the BNC, and it seems hard to construct acceptable ones. In particular, they do not work in the contexts identified by Houser (2010) where stative *do so* becomes much more acceptable since VPE cannot occur, for instance when the anaphor is non-finite (88):

- (88) ...or how Metabolife could own the copyright on an interview conducted by someone else. To do so/#do it, he said, would violate attorney-client privilege. (Houser (2010), ex. 34m p. 50)

These preferences reflect the semantics of main verb *do*, which is itself not a stative verb but a generic action verb. Halliday's pseudo-cleft test discussed above as well as the data from interrogatives with *do*, which all show reduced acceptability with states or non-agentives, suggest that these kinds of referent are less compatible with the sense of *do*. Additionally, the agentive meaning corresponds to the typical interpretation of *do* NP constructions such as *do business*, *do research* (excluding idioms such as *do the trick*, *do justice* etc.).

3.3 Apparent problems for the compositional analysis

The data reviewed in the previous section highlight the independence of the verb and pronoun components in *do it* anaphora, showing that *it* functions independently as a pro-VP, and that semantic restrictions on possible

antecedents are partly attributable to the semantics of main verb *do* and specifically non-idiom *do* NP constructions. Therefore there are good reasons for adopting a compositional analysis of *do it* and the pair *do this/that*.

Under this approach, it may be argued that what is traditionally described as the antecedent of the VP anaphor is really the antecedent of the pronoun, which is combined with the sense of main verb *do* to obtain the interpretation of the entire VPA. This antecedent could be expressed in the manner suggested by Cornish in terms of ‘acts’ or actions. This reflects the usual interpretation of direct objects of *do*, which may either directly denote actions (e.g., *do business, do research, do battle...*) or refer to actions performed on or with the object (e.g., *do the puzzle, do the dishes*).

The main challenge for this analysis comes, as noted, from the impossibility of restoring the VP antecedent under any form, since main verb *do* cannot take VP complements. The alternative suggestion that the antecedent is an action-denoting noun phrase, as proposed by [Simner et al.](#) (e.g., *the borrowing of a puzzle*) poses a similar problem as it is unnatural as the complement of *do*.

The next section offers a possible solution to these problems drawing on [Cornish \(1999\)](#)’s concept of ‘antecedent-trigger’, a linguistic element through which the actual antecedent, and thus the referent, is accessed. We will see that evidence of a semantic relation between *it* and the VP is actually a sufficient condition to consider that *it* and the VP are anaphorically related. Further, the problem of restoring the antecedent also comes up with other uses of *it* outside of VPA with *do*.

3.3.1 **Do* + VP and the antecedent of *it*

A paraphrase along the lines of that proposed by Cornish for *it* (‘the action of ...’) seems an adequate description of the antecedent, as shown in (89b). However, if we try to replace *it* by the VP antecedent, none of the possible VP forms are available, and the resulting sentence (89c) is ungrammatical:

3.3. APPARENT PROBLEMS FOR THE COMPOSITIONAL ANALYSIS

- (89) a. John passed his exam, but he only did it with difficulty.
b. *Semantics*: He carried out the action of passing the exam with difficulty.
c. *Syntax*: *He did passing/(to) pass the exam with difficulty.

These facts are somewhat striking seeing as such forms can occur with a variety of other verbs, as shown below in (90). The gerund, in particular, is externally an NP, and therefore of the right category for the object of *do*. The restriction in this case is therefore all the more surprising :

- (90) a. He contemplated/considered passing the exam.
b. He started revising/to revise for the exam.
c. I helped him revise for his exam.

Since there is no general exclusion against non-finite VPs as complements, the constraint we are dealing with is presumably specific to main verb *do*. Still, since VPs denote actions, it is unclear why they are unacceptable with *do*. Part of the problem is that the *do* + VP pattern is redundant, as as one would not expect there to be any significant difference in semantics between the VP alone and the putative **He did passing/(to) pass*. But this still does not explain why the sentence is ungrammatical and not just unnatural, as would be expected if this were a simple redundancy effect.

3.3.2 Antecedent and antecedent-trigger

Cornish (1992, 1999) argues that anaphora is not an independent grammatical relation between two linguistic elements (anaphor and antecedent) but rather a discourse process defined as a set of procedures for maintaining attention on entities already in high focus. Under this view, the term ‘antecedent’ as it is commonly used actually has two (divergent) senses: on the one hand, the conceptual representation allowing the addressee to locate the referent of a pronoun and, on the other, the linguistic expression

that gives rise to that conceptual representation. Cornish proposes to capture this distinction through the use of two different terms, ‘antecedent’ referring to the discourse model representation of an entity, and ‘antecedent-trigger’ being the linguistic expression through which this representation is accessed. In other words, while interpreting anaphoric expressions always involves accessing an antecedent, in the conceptual sense, this is not necessarily achieved thanks to an antecedent-trigger. Exophora, under this analysis, is characterised as the absence of an antecedent-trigger, rather than an antecedent.

Given the function of the antecedent-trigger, there is no requirement that it should be able to ‘replace’ the anaphor in the structure, although this would indeed show that the two are anaphorically related. In the case at issue, the VP associated with *do it* is simply the antecedent-trigger of *it*, while the VP is one of several possible ways of expressing the antecedent. The ungrammaticality of **do + VP* is therefore not an argument against treating *it* in *do it* as a VP anaphor. This problem is not unique to VPA, as there are further cases where replacing the anaphor by what is unambiguously its antecedent, in usual terms, leads to ungrammaticality. An example of this is where *it* takes a clausal antecedent (pro-S) but occurs in an environment where a finite *that-S* is excluded, for instance in SAI (91) or as the complement of a preposition (92):

- (91) a. A–I missed my train.
 B–How did it happen?
 b. *How did [_S (that) you miss your train] happen?
- (92) a. I will come, you can depend on it.
 b. *You can depend on [_S (that) I will come].

In both cases there is no reason to think the S is not the antecedent of *it*, even if it cannot replace the pronoun in the structure. However the restrictions here more general one, whereas the problem of VP antecedents

for *it* is specific to main verb *do*.

The concept of antecedent-trigger provides a way of dealing with such cases by correctly identifying the S as the element which helps interpret the anaphor, and not its structural counterpart. More generally, this approach makes it possible to maintain the compositional analysis of *do it* (as well as *do this/do that*, to which the above discussion also applies) while avoiding the independent problems raised by an analysis assuming a strict syntactic relationship between anaphor and antecedent. What is left to explain, under this approach, is what kinds of antecedents are accessed by pro-VP *it* as the complement of *do*, both in terms of their semantic properties and of their discourse status as regards saliency at the point where the anaphor occurs. The concluding section discusses the perspectives offered by the present analysis to address these issues.

3.4 Conclusion

Assuming a compositional structure for *do it* provides a good starting point for further analysis of its anaphoric properties. Separating *do* and the pro-form correctly predicts that alternation with *this/that* is possible, and reflects the fact that it operates under much the same conditions as when these pronouns are used outside of VP anaphora, e.g., with any other transitive verb or as sentence subject. In short, the usual properties of *this/that* play a role in the choice of *do it* vs *do this/do that*. These facts are not expected if one assumes that *do it* is non-compositional and/or has no internal structure.

The proposed analysis makes it possible to analyse separately the contribution of the main verb and pronoun to the anaphoric properties of the whole *do it* expression. It has been argued that *it* functions here as a pro-VP, as it is interpreted based on a previously-occurring VP. This description distinguishes the VPA reading of *do it* from the NPA reading

discussed by Simner et al, where *it* is a pro-NP anaphoric object of *do*. It is argued that both share the same structure *do* + pronoun, differing only with respect to the anaphoric function of *it*. This dual status of *it* is shown by the data from pseudo-clefting and passivisation, which supply evidence for syntactic and semantic identification between *it* and either an NP or VP. The pro-VP status of *it* is not unique to *do it* but exists independently of VPA, as *it* allows any kind of phrasal category as its antecedent. On a semantic level, restrictions on the antecedent of VPA *do it* can, under this approach, be redefined as constraints on the types of antecedents that can be accessed by *it* when it occurs as the complement of *do*. The antecedent should be actional and not stative; it should preferably denote an agentive action (in the strong sense of controlled and volitional, as opposed to e.g., natural/ accidental causation of an action), and finally it should be sufficiently salient to be accessed by *it*. In short, restrictions on the types of referents for *do it* are driven by the semantics of main verb *do* on the one hand, and the anaphoric properties of *it* on the other.

The main challenge for the analysis was to explain why *it* and not *do it* is the anaphoric element when the VP antecedent cannot be restored as the complement of *do*, due to the ungrammaticality of **do* + VP. The solution proposed relied on Cornish's divide between antecedent/antecedent-trigger, a refinement of the traditional concept of 'antecedent' that distinguishes the formal and semantic aspects of anaphora. Following this approach, the VP antecedent of *it* is instead characterised as its antecedent-trigger, a linguistic expression through which an appropriate antecedent, or representation in the discourse model, is accessed. Since the antecedent-trigger is merely a 'pointer' to the sense of the antecedent, it is optional and need not be structurally identical to a full phrase that could replace the anaphor. There is also no requirement that it should be able to replace the anaphor in some form or other, even though this is often the case. For this reason the ungrammaticality of **do* + gerund was treated as a separate

problem which does not call into question the anaphoric relation between *it* and the VP. Other examples of the ‘replacement’ problem with *it* were discussed for cases where the relation between *it* and the antecedent is well-established and beyond doubt, for instance in SAI with pro-S *it* (*how did it/*that + S happen?*). The approach adopted here acknowledges these independent problems while capturing the fact that the pronoun is interpreted based on the meaning of the VP/S.

Another advantage of the analysis is that it can be extended to *do this/do that* which share the same structure *do + pronoun*. As with *do it*, possible antecedents for these two forms are those which can be accessed by *this/that* when used as complements of main verb *do*. Further research must determine exactly what kind of antecedents are possible, and what degree of saliency they usually have in the context where they occur. From [Miller \(2011\)](#)’s observations it appears that in this respect the properties of *this/that* only partially map onto *do this/do that*, suggesting the saliency conditions differ when the antecedent is an action. Finally, treating *do it* and *do this/do that* as compositional can explain a number of important differences with *do so*, which seems not to be compositional syntactically (it is not straightforwardly analysable as *do + so*) and semantically (it allows stative antecedents under certain conditions).

Providing an account of the structure of *do it* is only the first step in explaining its anaphoric properties. The rest of the analysis must describe how semantic restrictions on the antecedent and saliency requirements can be explained based on the structure proposed here. This involves determining which actions can appropriately be characterised as ‘doing something’, and therefore be expressed anaphorically as complements of *do*. Comparison with NPA *do it* and *do NP* collocations (*do the math/dishes, do the V-ing*) should provide insights into this question. The analysis can also be extended to exophoric *do it* (and *do this/do that*), where no antecedent-trigger is present and the anaphor is resolved largely based on the extralin-

CHAPTER 3. COMPOSITIONALITY

guistic context. Finally, assuming that anaphoric *do it* is compositional distinguishes it from idiomatic uses such as *that did it* (= ‘that worked’, ‘that did the trick’) where *do it* is non-compositional.

Chapter 4

Agentivity

4.1 Introduction

Various authors have claimed that VPAs require, or at least prefer, an agentive antecedent, by which is meant that they should denote actions under the control of human agents. The exact statement of this constraint varies: some regard it as being truly a condition on the use of VPAs (Souesme, 1985; Huddleston and Pullum, 2002; Culicover and Jackendoff, 2005), while for others it is merely a semantic preference which to a certain extent admits of exceptions (Quirk et al., 1985; Miller, 2011). Such accounts also differ as to whether the agentivity requirement, or preference, is supposed to apply ‘across the board’ to all VPAs including *do so* (Culicover and Jackendoff, 2005) or to some of them only. The notion of an agentivity requirement is also sometimes intended to rule out stative antecedents, which have usually been deemed impossible (though there is now ample evidence that *do so* does in fact allow them, especially in its non-finite uses, see for instance Michiels 1978; Houser 2010; Miller 2013).

Whatever the differences between them, most of the above analyses rely exclusively on constructed examples, with the exception of corpus-based studies such as Souesme (1985) and Miller (2011). Whether or not

they are considered acceptable, discussion of non-agentive VPAs in the works cited tends to be limited, and there have been no studies focusing on the alternation between non-agentive *do it/this/that* on the basis of attested data. This chapter will therefore reconsider the agentivity requirement or preference ascribed to *do it/this/that* in the light of extensive data from BNC. We begin by reviewing the various conceptions of agentivity in general and as understood by those authors who mention it as a semantic constraint on *do it/this/that*. We then discuss preliminary data on agentivity and VPAs based on the 900-occurrence sample from the BNC, which contains few non-agentive examples (see Chapter 2, section 2.3.14), before presenting the results of a separate corpus study in which specific queries were used to search the BNC and retrieve more non-agentive occurrences than in the initial sample.

4.2 Agentivity constraints on VPA antecedents

The idea of an agentivity constraint on the antecedent is encountered in some of the earliest accounts of VPA, such as Quirk et al. (1985) and Souesme (1985). For Quirk et al., *do it* differs from *do that* ‘for some speakers’ in that it requires an antecedent that conveys ‘volition on the part of the subject’ (p. 877). The examples provided are reproduced below (it is not specified whether *do this* is also subject to this constraint). According to Quirk et al., *do that* is preferred over *do it* in (93), where the antecedent is non-agentive, but *do it* becomes acceptable if B replies by asking about the manner of the action (94) rather than the cause as in (93). They argue that in (94), the subject is in fact viewed as agentive, more precisely, as they put it, ‘B talks as if the chicken were still alive and had control over its movements’ (p. 877), an interpretation which is facilitated by *how*.

- (93) A: When you chop off a chicken’s head and it’s already dead, it still kicks a few times. (p. 877)

4.2. AGENTIVITY CONSTRAINTS ON VPA ANTECEDENTS

B: Why does it do that/*it?

(94) A: When you chop off a chicken's head and it's already dead, it still kicks a few times.

B: I wonder how it does it. (p. 877)

In spite of these data, Quirk et al. still maintain that non-agentive *do it* is 'abnormal' on the grounds that 'transitive *do* is both (i) dynamic and (ii) agentive, i.e, it refers to some action that is voluntarily performed by the subject' (p. 878). This rules out stative antecedents (95) and various non-agent subjects such as experiencers (96).

(95) They think he is mad.

*We do it too. (Quirk et al. (1985, ex. 1, p. 878))

(96) A: Bob might have heard the strange noises.

B: (?) He might well have done it. (ex. 2, p. 878)

Quirk et al. (1985) argue that although there is 'divided usage' regarding non-agentive *do so* (an agentive antecedent being preferred in American English), it is still more acceptable than *do it* or *do that* in the above examples.

Souesme (1985) is more radical and claims, on the basis of a corpus of approximately 50 occurrences, that *do it* (as well as *do what* or *do something*) requires an agentive antecedent, which he defines as 'a conscious and voluntary activity carried out by the subject' (*une activité consciente et volontaire du C₀*). He identifies the sentences in (97)-(98) as potential counter-examples, but goes on to explain that they are in fact agentive:

(97) Clearly Eliza will not pass as a duchess yet; and Higgins's bet remains unwon. But the six months are not yet exhausted; and just in time Eliza does actually pass as a princess. For a glimpse of how she did it imagine an Embassy in London one summer evening after dark. (ex. 103, p. 41)

- (98) She was wearing jeans, which I regretted, but you can't stop them doing it. (ex. 104, p. 41)

That (98) is agentive is unquestionable: it seems odd to suggest, as Souesme does, that *wearing jeans* could in any way be interpreted as non-agentive. The case of (97), on the other hand, calls for further discussion. If *pass as a princess* here is understood as giving a certain impression, or appearing to have a certain rank in society, then clearly the antecedent is not agentive insofar as it is not entirely under the subject's control. However, if we follow Souesme's criteria for the agentivity of *do it* antecedents, it also cannot be considered as merely involuntary or 'unconscious', as it is at least the result of Eliza's efforts to give a particular image of herself in public. Souesme explains that the antecedent describes an 'attitude', but *how* entails volition on the part of the subject. This is probably a step too far: *how* is certainly compatible with non-agentive predicates (cf. *How did you get ill?*), but it is arguable that the use of a manner expression highlights the subject's involvement in the process, which in turn may facilitate (although not necessarily result in) an agentive interpretation.

It follows from the proposed requirement of an agentive antecedent that the subject must be animate. Souesme nevertheless cites one example with an inanimate subject, shown in (99). The use of *do it* here, he argues, is justified by the fact that dioxin is the 'agent' or 'trigger' (*déclencheur*) of a chemical process. However, this description involves a different use of the term 'agent', which is no longer limited to animate volitional agents as initially suggested.

- (99) But the best way to resolve what dioxin does and how it does it may be to focus on a much smaller target—the cell. (ex. 106, p. 43)

It must be pointed out that this is not, strictly speaking, a case of VP anaphora, since the antecedent-trigger here is the NP *what dioxin does*, which is directly substitutable for the pronoun *it*. The example does how-

4.2. AGENTIVITY CONSTRAINTS ON VPA ANTECEDENTS

ever suggest that causal processes are at least one situation in which non-agentive *do it* may be acceptable. Souesme's position on *do this/that* with regards to agentivity is unclear, as he does not mention them in his discussion of *do it* (there is also no mention of *do so*). Presumably, since they are also *do NP* constructions analogous to *do it* or *do something*, they are subject to the same semantic constraints.

Huddleston and Pullum (2002) follow Quirk et al. (1985) in claiming that *do so* allows non-agentive uses (cf. 100), but they consider that both *do it* and *do that* must have an agentive antecedent (presumably this also applies to *do this*, though they do not explicitly say so). Culicover and Jackendoff (2005) go further and state that all VPAs including *do so* must be agentive. The examples given in support of the agentivity constraint are all of the non-causal type, similar to Quirk et al.'s (96) above. Culicover and Jackendoff call this category 'non-action events' (101), as opposed to an agentive action or a state. They only provide examples with *do so*, such as (101), but their analysis implies that the *do it/this/that* variants given in (102) should be just as unacceptable as the original sentence.

(100) When the tree fell, it did so/#did it with a loud crash. (Huddleston and Pullum, 2002, ex. 49, p. 1532)

(101) ?* Robin fell out the window, but Leslie didn't do so. [Non-action event] (Culicover and Jackendoff, 2005, ex. 2b, p. 284)

(102) ?* Robin fell out the window, but Leslie didn't do it/this/that.

The view that *do so* must be agentive is a long-standing one, proposed as early as Lakoff and Ross (1976), but it has been questioned by Michiels (1978); Houser (2010); Miller (2013). Houser provides both corpus-based and experimental evidence that *do so* in fact allows both non-agentive and stative antecedents (the latter especially when *do so* is non-finite) but considers that the behaviour of *do it/this/that* in this respect remains an open question. Later experimental work by Miller (2013) on stative *do*

so reaches similar conclusions to Houser's regarding the acceptability of stative antecedents, but does not address the agentivity facts in detail.

In a short corpus-based study of the factors influencing VPA choice, Miller (2011) observes that non-agentive occurrences of *do it/this/that* are rare but attested, but no stative examples are found in his corpus. He provides the following case of non-agentive *do it* where the antecedent is causal, as in Souesme's examples above:

- (103) Although straight narrative can advance a plot faster and less circuitously, dialogue often does it more interestingly. (ex. 1b, p. 83)

The existence of such examples casts doubt on the idea of an agentivity requirement for *do it*. Attested cases with *do this/that*, such as (104)–(105) below, further call into question the existence of such a requirement on VPAs generally:

- (104) Betaadrenergic agonists elevate the er stimulate chloride secretion into the lumen, and the way they do this is they raise cyclic AMP levels obviously and activate the chloride channels (J8K)
- (105) As the Right Honourable Gentleman is aware, our top priority is to get inflation down, and I would also hope that he would be aware that the aim of doing that in the longer term has to be by making the price of money more expensive. (KRT)

Although the vast majority of occurrences in the corpus are agentive, the above sentences are at least indicative of the possibility for all *do* + NP VPAs to have a non-agentive antecedent at least in some cases. As pointed out earlier, what the attested examples examined so far have in common is that the subject causes the action rather than simply undergoing, say, a change of position, as the falling tree in (100), or experiencing a stimulus as in (96). For instance, in (104) the subject triggers a chemical process and is therefore the cause (or agent, in a loose sense), of that process,

4.2. AGENTIVITY CONSTRAINTS ON VPA ANTECEDENTS

much as in Souesme's example. In (105), *get inflation down* means 'cause inflation to go down'; it is a process the speaker hopes to bring about, but he does not have direct control over it. More such examples are shown in (106)–(108). They are intuitively more natural-sounding than the invented examples from Culicover and Jackendoff and others, suggesting that semantic constraints on VPAs may be better expressed in terms of causality, rather than agentivity in the intentional sense of the word.

- (106) [T]he fact is that the reason why you, why the menopause occurs is because the ovaries stop functioning, they stop (pause) producing oestrogen (pause) and every woman's ovaries does [*sic.*] this and they do it (pause) you know, at all varying ages (pause) the average age is fifty. (FL4)
- (107) There is, indeed, a snake—the Formosan banded krait—that gives you a type of instant miocencia, and the discovery that the venom of this animal contains a toxin that can do this has, indirectly, led to the elucidation of the mechanism behind miocencia gravis. (KRF)
- (108) The name of the game is how can we reduce the amount of offending in this country. If cautioning could do that we'd all be right behind it. (HE5)

The possibility for VPAs to express non-volitional causation is in fact compatible with the semantics of main verb *do*, which is itself not limited to agentive situations, and can readily express simple causation, as in *This cream will do wonders for your skin*, or the following attested case found in the BNC:

- (109) The religious wars of the later sixteenth century, between Catholics and Protestants, did much damage to people and to places in the Pyrenean region, Pau included, and Jeanne d'Albret's army commander, Montgomery, who recaptured Pau from the troops of King

CHAPTER 4. AGENTIVITY

Charles IX, achieved an especially nasty reputation for the reprisals and pillagings he carried out. (FA2)

However, it is also possible to find rare instances of antecedents that lack the causal property. Below is an example with *do it*:

- (110) ‘I’m afraid of slipping in the dark and cracking my head on the edge of the bunker,’ I had said. ‘I did it once and it has made me afraid.’ (AC7)

Examples of this type pose a greater challenge for the analysis than the causal non-intentional cases, since they do not involve either volition or causality, raising the question of what the semantic restriction on VPAs might be. There is also no expectation that non-idiomatic constructions involving main verb *do* could have a subject that is not an agent or a cause. The greater acceptability of *do it* here compared to similar constructed examples from [Culicover and Jackendoff](#) and others may be linked to the presence of the adjunct *once*. Note that a paraphrase with *happen* (*It happened to me once*) is possible here.

In short, the requirement of an agentive antecedent for VPAs seems questionable. The constraint against stative antecedents, on the other hand, seems a much stronger criterion, as no examples were found in the BNC sample (note, though, that four examples with stative antecedents were found in the extended sample described in [4.4](#) and are given in [4.6.9](#)). *Do it/this/that* also do not appear to be a felicitous alternative to *do so* in attested stative examples such as the following from [Houser \(2010\)](#):

- (111) The six genes that have the largest t-statistics do so (#do it / this / that) by virtue of having denominators close to zero, implying near constant expression levels. (ex. 1a, p. 62)

4.3 Defining agentivity

Before a more in-depth analysis of the corpus data can be undertaken, it is necessary to clarify what is meant by agentive. We have seen the term used in various ways in relation to VPAs: while most authors appear to restrict the category of agent to animates, [Souesme](#) implies that inanimate causes may also count as agents, though he argues that animate human agents are preferred with *do it*. Part of the problem is therefore to determine which semantic features are required to consider an antecedent as agentive. A further question is how the presence of such features might be established in the first place. [Cruse \(1973\)](#) offers an interesting discussion of these issues, which we review in the next section.

4.3.1 Agents and non-agents

The two main definitions of agentivity cited by [Cruse](#) diverge on the status of inanimate agents of the type just discussed. According to [Fillmore \(1968, p. 24\)](#), agentive is simply ‘the case of the typically animate perceived instigator of the action identified by the verb’, leaving open the possibility of inanimate agents (although these would be atypical, in his conception). There is also no explicit requirement that animate agents should be volitional, though the term ‘instigator’ would suggest that this is the default situation. [Gruber \(1967, p. 43\)](#), on the other hand, is more categorical, stating that the subject of agentive verbs ‘refers to an animate object which is thought of as the wilful source or agent of the activity described in the sentence’, therefore treating Fillmore’s typical properties as actual constraints. If we follow the first of these definitions, then all non-agentive occurrences in our corpus (with the exception of non-causal situations) are actually a less typical kind of agentive predicates. On the other hand, if we retain [Gruber](#)’s criteria, then the non-agentive set is clearly separate from the rest.

[Cruse](#) argues that Fillmore's animacy requirement is questionable, seeing as various kinds of inanimate causing forces are also referred to as 'agents'; these include so-called 'natural agents' such as the sun, wind and so forth. [Gruber](#)'s definition, in his opinion, is vaguer than Fillmore's, as the difference between 'source' and 'agent' is left unexplained.

An alternative approach mentioned by [Cruse](#) is to view agentivity as a relation holding between the verb and its subject, as proposed by [Halliday \(1967\)](#), in other words, in terms of the subject's involvement in a particular action (whether or not it is causal, volitional, and so on). This is in line with the usual conception of 'agent' as a semantic role.

4.3.2 Linguistic tests for agentivity

Aside from real-world definitions of agentivity, [Cruse](#) reviews a series of linguistic tests that may be used to characterise a VP or its subject as agentive. [Gruber \(1967\)](#) proposes several of them, among which the possibility of substituting the VP 'in all circumstances by the phrase *do something*', or the possibility of modification by manner adverbs such as *carefully* or purpose adjuncts such as *in order to*. The last two of those predict that *jumped down the stairs* below is agentive while *fell down the stairs* is not.

(112) John jumped down the stairs deliberately/in order to impress his friends.

(113) #John fell down the stairs deliberately/in order to impress his friends.

The test may be extended to show that predicates of the type in (113) preferably combine with adjuncts or adverbs that explicitly deny volition or control on the subject's part:

(114) John fell down the stairs accidentally/through no fault of his own.

The *do something* test, on the other hand, is not viable as it stands, as shown by the following pair from [Cruse](#), in which the oddity of (116) (for

real-world rather than linguistic reasons) does not change the fact that *killed his father* is agentive.

(115) John killed his father. That's why he was hanged. (ex. 3, p. 13)

(116) John did something. That's why he was hanged (ex. 4, p. 13)

It makes more sense to think of the test in terms of entailments, rather than strict replaceability, as [Cruse](#) suggests later on. For instance, (115) is agentive since 'John killed his father' entails 'John did something'. By contrast, *broke* in (117) does not carry a similar entailment, and is therefore not agentive:

(117) ?The vase broke: therefore it did something. (ex. 10, p. 14)

A more precise implementation of the test cited by [Cruse](#) comes from [Halliday \(1967\)](#), who relies on pseudo-clefts of the type *What X did was VP* to test for relationship with *do*. For instance *What John did was punch Bill* is more natural than *?What the vase did was break*, suggesting the former is agentive and the latter is not. Note that causal processes also pass the test, as the following examples from [Cruse](#) illustrate:

(118) What the computer is doing is calculating the correlation coefficient. (ex. 15, p. 16)

(119) What the bullet did was smash John's collar bone. (ex. 18, p. 16)

As [Cruse](#) admits, though, the function of the *do*-test is ultimately to verify the compatibility of some action with the meaning of main verb *do*; it is only indicative of agentivity if main verb *do* itself is considered agentive, a property which remains to be defined at this stage.

4.3.3 Semantic features of agentivity

In order to escape the circularity of the *do*-test in its various forms (the result depends on a property of the VP that the test itself is supposed to

determine), [Cruse](#) identifies a set of four recurring semantic features of VPs that pass the test: volitive, effective, initiative and agentive. ‘Volitive’ applies wherever the subject is a volitional agent (e.g., *What John did was not eat anything for two days*); it corresponds to agentivity in the narrow sense of volition as conceived of by [Gruber](#). ‘Effective’ describes an object exerting a force by virtue of its position, motion or other. This category includes inanimate causes as well as certain kinds of states:

(120) The flying stone broke the window (ex. 53, p. 19)

(121) These columns support the weight of the pediment. (ex. 50, p. 19)

A further distinction can be made between the two ‘effective’ examples above. In (120) the subject is causal, whereas (121) does not clearly involve causality, but rather ‘passive’ exertion of force.

The ‘agentive’ feature is present wherever an entity is ‘regarded as using its own energy in carrying out the action’ (p. 21). The term covers animates as well as inanimates, specifically including ‘living things, certain types of machine, and natural agents’ (p. 21). If the subject is animate, this feature appears to imply volition, with inanimates, control is the defining property of agentives:

(122) John moved (himself) to avoid the falling stones. (ex. 67, p. 21)

(123) The machine automatically switches (itself) off at 6 p.m. (ex. 68, p. 21)

The distinction between this feature and the ‘effective’ one implies that causality is a possible but not a required property of agents. For instance, the continuation in (a) below illustrates the effective case, while (b) is agentive:

(124) John smashed the window

(a) when he fell against it.

(b) with a stone. (ex. 86, p. 22)

However, it is unclear how (b) differs from the volitive situation, if at all, since it is obvious that a deliberate action is being described. It may also be pointed out that the agent does not use only his own energy in carrying out the action, since the stone is ultimately the cause of the window breaking (compare *John broke the window with his fist*).

A fourth feature, called ‘initiative’, which is present in situations where the subject controls the actions of another entity, e.g., *The warder marched the prisoners across the yard*, is of little interest here, as it seems to be analysable as a subcase of volitive (or possibly agentive), rather than a separate category in its own right.

We are now in a position to characterise agentivity in a more precise manner than was initially the case. Cruse’s analysis leads us to identify the following three features as essential to the classification of predicates as agentive or non-agentive: animacy, volition, and causality. These may be represented schematically as follows: +/– animate, +/– volitional and +/– causal. Following Gruber, we restrict the category of agents to animates, subject to the criterion of volition. Entities lacking any of the three features are consequently non-agentive. Although, as seen above, various kinds of inanimates are also sometimes described as agents, the proposed definition allows us to clearly distinguish between ‘true’ agents and inanimate non-agents. It is also consistent with the conception of agentivity which is implicitly adopted in most accounts of VPA which regard it as a necessary or at least typical property of their antecedents.

It follows from the above definition that there are several types of non-agentive situations (some of these can still be considered actions, others may be referred to as processes). The three features give us eight theoretical combinations in total, of which two are inherently contradictory (–animate, +volitional, +/–causal). It is also unclear whether an animate subject can be volitional without also being causal (perhaps this might apply some of Cruse’s ‘effective’ cases). The list below shows the remaining

CHAPTER 4. AGENTIVITY

five possible combinations, and indicates (roughly) how they map onto different semantic roles for the subject:

1. + animate, + volitional, + causal: agents
John picked up his keys
2. + animate, – volitional, + causal: involuntary animate causes
John sneezed
3. + animate, – volitional, – causal: animate themes, patients, recipients or experiencers:
Mary broke a leg/saw a ghost/received a parcel
4. – animate, – volitional, + causal: inanimate causes
The fire destroyed the building
5. – animate, – volitional, – causal: inanimate patients or themes:
The tree fell during the storm/the window broke

Although this classification leaves us with four main types of non-agentive antecedents, it is apparent that further refinements are possible; for instance, type (3) covers quite different situations, and we will see later on that it is also necessary to distinguish between several subcategories of inanimate causes with type (4). Preliminary data from the BNC show that antecedents of all types except (5) are attested. In the rest of this chapter we will seek to determine how frequent the different non-agentive types are with *do it/this/that*, and whether cases of type (5) may in fact be acceptable.

4.4 Non-agentive VPAs: a corpus study

4.4.1 Overview of the BNC data

A brief survey of the corpus confirms that agentive antecedents are very much the typical case with VPAs: only 20 of the 900 tokens (just a little over 2% of the data) are non-agentive. With *do it*, the figure is as low as 4 of 500 occurrences, less than 0.8% of tokens. Non-agentive *do this/that* appear to be more common, with respectively 3.5% and 4.5% of non-agentive antecedents, but the difference is much too small to confirm this hypothesis at present.

Although relatively small, this preliminary corpus is semantically diverse. Causal subjects (125) are the typical case (15 occurrences), but there are also a few examples of theme (128) and even recipient and patient subjects (126)-(129):

- (125) [*Documentary: working conditions at an industrial bakery*]
— Here we have Brian using a depositor er to get this batter into the pans. You may think this is a productivity issue but in fact erm our purpose here is to save stress on the individual.
— And does it actually do that?
— Yes I think that it does. (K6B)
- (126) Recently somebody has won just over nine hundred thousand pounds, and it's interesting to note that he did this without exercising too much skill, as I'm sure he'd be the first to admit, because he enters the same numbers every week. (KRF)
- (127) First of all the British Met Office erm receives information from all over the world, and it is able to do this because it is part of a massive weather organisation called the World Met Organisation, and this is linked by very, very high-speed communications worldwide. (KRH)

CHAPTER 4. AGENTIVITY

- (128) ‘I’m afraid of slipping in the dark and cracking my head on the edge of the bunker,’ I had said. ‘I did it once and it has made me afraid. (AC7)
- (129) [I]f I broke a leg I’d have done it whether you were with me or not. (APU)

Of these, only the *do it* sentences seem to show reduced acceptability. (129) may also appear less natural-sounding due to the (possibly non-standard) use of the simple past rather than the past perfect *if I had broken my leg*. In (128), the presence of the adjunct *once* makes *do it* more acceptable than in a polar question such as (125), but a variant with *happen* (*It happened to me once*) still sounds more natural.

The *do this/that* examples are more natural-sounding, but note that *do it* also seems acceptable in (126) (probably due to the influence of *win*, which can be related to the ‘success’ meaning of idiomatic *do it*). It is much less natural in (125) (*#And does it actually do it?*), but this is partly because the VPA occurs in a polar question and is separated by a speech turn from the antecedent-trigger.

Within the causal category, there is variation in the types of antecedents encountered, for instance, inanimate causes include (amongst others) biological or chemical processes as in (130) (Cruse refers to these as ‘natural agents’) as well as more abstract cause-effect relations (131).

- (130) There is, indeed, a snake – the Formosan banded krait – that gives you a type of instant miocencia, and the discovery that the venom of this animal contains a toxin that can do this has, indirectly, led to the elucidation of the mechanism behind miocencia gravis. (KRF)
- (131) Our primary objective is to prevent, or reduce, re-offending. Now we know that our diversion system allows us to do that. (HE5)

Preliminary data reveal interesting semantic discrepancies with respect

to non-agentive antecedents. The *do that* set has only causal antecedents, while *do this* is also encountered with theme or recipient subjects. With *do that*, there is a tendency for the antecedent to express a situation over which the subject has no direct control but which is the intended result of his/her actions. Thus the subject may be regarded as causal in the sense of bringing about a particular situation. (131) is an example of such an antecedent, since it is impossible to directly reduce re-offending, but measures can be taken to avoid it. The subject of *do that* in this usage may also refer to an instrument used to achieve a particular goal, as with *If cautioning could do that (= reduce the amount of offending)* in (108). As for *do it*, aside from the two examples already mentioned, the remaining two are causal, but have either *do this* or *do that* as the antecedent-trigger:

- (132) — And you need to access, you need to key in to certain areas as well, so television could be the ideal way to do that.
— But it's not doing it at the moment. (FLR)

This preliminary sample, although relatively small, does at least give us a more precise idea of the non-agentive use of *do it/this/that*. The most important finding is that they do not in fact require an agentive antecedent, contrary to the claim often made in the literature. Instead, we can concur with the weaker statement of the agentivity constraint, which merely regards it as the preferred usage of VPAs. It was also noted that the premise that main verb *do* is always agentive is in fact wrong, as it can readily express causation with animate or inanimate subjects. It is therefore not surprising that with most occurrences found so far the subject is a cause of the action in various ways—an animate ‘involuntary’ cause, a force of nature, a material entity, or possibly an event or situation. However, we also find antecedents where the subject is neither a cause nor a volitional agent. There is no evidence so far to suggest that non-agentive VPAs are less acceptable than agentive ones, though it appears that *do it* more strongly prefers agentive antecedents than *do this* or *do that*, and especially resists

non-causal ones. As argued, however, at least one of the attested examples is degraded partly for reasons independent of agentivity.

As already pointed out in section 4.2, stative VPAs are not attested in the corpus. This is unsurprising considering that main verb *do* itself non-stative. Moreover, the exclusion of stative antecedents is also a more consensual matter in the literature, as it has never been suggested that they are in fact possible.

Given the very limited data on non-agentive uses in the current corpus, the conclusions sketched above are only tentative, and must be reconsidered in the light of more extensive corpus analysis. In what follows we discuss the difficulties of such an approach in view of the limitations of the BNC, as well as methods employed to retrieve a larger number of non-agentive tokens.

4.4.2 Data collection methods

The constitution of a corpus of non-agentive VPAs faces two main hurdles. First, the BNC is limited to part-of-speech tagging and does not encode any semantic features, making it impossible to search for agentive predicates. Since, moreover, anaphoric relations are also absent from the corpus tagging, there is in fact no way to search for antecedent-triggers as such, whatever their semantic or syntactic properties. A further difficulty is that there are no recurring grammatical features of non-agentive VPs, as most often the agentive or non-agentive meaning is determined by the head verb. Although as discussed earlier, certain modifiers can impose a non-agentive reading on the VP, they are by no means a condition for this interpretation to arise.

These limitations seem to leave no option but to simply search for VPA occurrences and determine agentivity on a case-by-case basis, but as we saw from the data reported in (4.4.1), this approach is much too broad to yield enough results. Fortunately, there are at least two possi-

ble workarounds for this problem, both of which rely on our definition of agentivity as being strictly a property of animates. We explain them briefly below and discuss technical details and results in the next section.

The first method is to search for occurrences of VPAs with inanimate pronouns as subjects, e.g., *it does it/this/that*. Since under the definition we have adopted, inanimates cannot be agents, the sentences obtained are highly likely to be non-agentive. From a strictly linguistic perspective, this is a fairly reliable method, even if we take into account the fact that inanimate *it* may also refer to animals or collective entities such as institutions, companies, or countries which are actually animate (examples of these with VPAs are attested, but not extremely common). There are, however, some disadvantages linked to the design of the BNC. Since it does not encode grammatical functions such as subject or object, the search has to be restricted to contexts where the pronoun immediately precedes the VPA to reduce the amount of irrelevant data. This is in addition to the fact that the BNC tags do not indicate phrasal category, and do not distinguish between main verb and auxiliary *do*, as already noted in Chapter 2.

An alternative approach consists in what may be called a ‘targeted text search’, which consists in simply searching for occurrences of VPAs—irrespective of syntactic context—in texts from specific domains that are more likely to make reference to inanimates, such as natural sciences, engineering, or computing. This method makes arguably loose assumptions on the content of text (though it is likely that references to inanimates are more frequent in such texts, they may not be significantly more frequent), and as such it is less precise than the first. However, it allows for a much wider range of syntactic contexts. In addition, although the intention is to retrieve inanimate subjects, the search also allows us to find animate non-agents as well. These will of course not be found by the first method.¹

1. Two further ways of looking for non-agentives initially considered were eventually rejected. One is to search for inanimate nouns, rather than pronouns (*the computer/machine did it; the rain/wind did it, etc.*). This is tedious, since one first has to

A total of 164 occurrences was obtained using a combination of these two methods (the majority of them using the first method). The corpus collected in this way was then annotated with the animacy, causality, and semantic role of the subject, along with finiteness of main verb *do*, presence of an adjunct and type of adjunct. Section (4.5) further explains the implementation and usefulness of each method. Coding methods and results of the semantic roles analysis are in (4.6) and (4.6.3) respectively.

4.5 Results: frequency of non-agentive VPAs

4.5.1 Method 1: pronominal subjects

A number of queries were used in addition to the one suggested above to make up for the syntactic limitation of method 1 by covering a greater variety of patterns. This was achieved by searching for various combinations of tenses and aspect² as well as contexts where the VPA is under the scope of a modal. The full range of queries is shown below:

```
it|this|that [do] it|this|that
```

searches for finite or non-finite uses ([do] stands for all forms of *do*) preceded by an inanimate pronoun (which, in some cases, is not itself the subject), e.g., *It does/did it/this/that*.

draw up a list of such nouns (which is somewhat open-ended, raising the question of what nouns to chose, and how) to avoid searching for them one at a time. It also suffers from the same limitations as method 1 in terms of syntactic context, and returns very few occurrences. Another method, as suggested earlier, is to search for adjuncts that impose a non-agentive reading on the VP, such as *accidentally*, *inadvertently* or perhaps *automatically*. This approach raises similar problems to the first and, like it, returns few results.

2. The present perfect/past perfect progressive combinations were excluded as the query returns only one occurrence, where the antecedent is agentive.

4.5. RESULTS: FREQUENCY OF NON-AGENTIVE VPAS

```
it|this|that _vm* do it|this|that
```

searches for occurrences of VPAs preceded by a modal auxiliary (*_vm**), such as *It can/could do it*.

```
it|this|that [have] done it|this|that
```

searches for occurrences with perfect aspect (*It has/had done it*)

```
it|this|that [be] doing it|this|that
```

searches for occurrences in the progressive.

Increasing the number of queries in this way has the disadvantages of also increasing the quantity of noise that has to be filtered out, since results include all instances where the forms being searched for occur in the specified order, irrespective of the syntactic relation (if any) between them. The above queries returned a total of 574 occurrences. Cases where the *do* + pronoun sequence is not a VPA were removed as well as those where the context is insufficiently clear to retrieve the antecedent. The remaining sample contains 187 occurrences of VPA, of which nearly half are non-agentive. Results appear in Table 4.1 below.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Agentive	26 (51%)	47 (58%)	24 (43.6%)	97 (51.9%)
Non-agentive	25 (49%)	34 (42%)	31 (56.4%)	90 (48.1%)
	51	81	55	187

Table 4.1: Agentive and non-agentive VPAs in the BNC: method 1, pronominal subjects

Contrary to what was observed for the initial corpus, *do it* is more frequent in non-agentive uses than *do this*, but it is less frequent than *do that*, which again shows the highest proportion of non-agentive antecedents.

However, there was no significant effect of agentivity ($p = 0.2$), and pairwise comparisons between VPAs also showed no significant differences (*do it/do this*: $p = 0.4$; *do it/do that*: $p = 0.4$, *do this/do that*: $p = 0.1$). The fact that such large ratios of non-agentive occurrences are encountered demonstrates the efficacy of the method used.

4.5.2 Method 2: STEM texts

The first method relied on syntactic strategies to retrieve occurrences of VPAs with inanimate subjects. The second was similarly designed to retrieve such occurrences, whilst also circumventing the syntactic limitations of the queries used the first method. It makes use of semantic heuristics by restricting the search to BNC sections belonging to domains that are more likely to make reference to inanimates—and thus to describe non-agentive actions—than the fiction or spoken texts used for the initial corpus, in which the range of topics is much wider.

The sub-corpus used for this search is comprised of a variety of both academic and non-academic texts covering such topics as the natural sciences, engineering and computer science, or (more rarely) physics and mathematics. With the exception of natural science lectures, all of the data consist solely of written discourse. The written portion also includes texts from more general sources whose main focus is not science, such as newspaper articles.

The full list of sections included into the search is shown below along with a brief description of each:

- *S_lect_natsci*: natural science lectures recorded at British universities
- *W_ac_nat_sci*: academic publications in natural science as well as astronomy, physics and mathematics

4.5. RESULTS: FREQUENCY OF NON-AGENTIVE VPAS

- *W_nonac_nat_science*: non-academic publications in natural science, articles from the *New Scientist*
- *W_ac_medicine*: academic publications in medicine, mostly in the form of journal articles
- *W_nonac_medicine*: non-academic medical publications, including self-help guides or health information leaflets
- *W_ac_engin*: academic publications in engineering, particularly electronics and computer science
- *W_nonac_engin*: computer science periodicals
- *W_newspaper_brdsh_t_nat_science*: science articles in national broadsheets
- *W_newspaper_other_science*: science articles in local newspapers

This method returned a total of 936 occurrences. 342 unclear and non-VPA examples were excluded, as well as a further 34 which had already been found with the first method. Of the remaining sample of 560 occurrences, 70 (12.5%) are non-agentive. Not surprisingly given the size of the sample and the queries used, the ratio of non-agentive uses is much lower than in the data obtained using the first method. However, it is still quite high compared to the 2% ratio found in the initial corpus. Results are shown in Table 4.2

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Agentive	153 (90.5%)	291 (85.3%)	46 (92%)	490 (87.5%)
Non-agentive	16 (9.5%)	50 (14.7%)	4 (8%)	70 (12.5%)
	169	341	50	560

Table 4.2: Agentive and non-agentive VPAs in the BNC: STEM texts

Contrary to the method 1 results, *do that* here is numerically the least frequent form in non-agentive uses, and *do this* the most frequent. However, there was no significant effect of agentivity ($p = 0.1$), and differences between VPAS were not significant (*do it/do this*: $p = 0.1$, *do it/do that*: $p = 0.7$, *do this/do that*: $p = 0.2$).

	Agentive	Non-agentive	TOTAL
W_non_ac_tech_engin	83 (82.2%)	18 (17.8%)	101
W_non_ac_nat_science	214 (84.6%)	39 (15.4%)	253
W_ac_nat_science	19 (86.4%)	3 (13.6%)	22
W_ac_tech_engin	38 (90.5%)	4 (9.5%)	42
S_lect_nat_science	14 (93.3%)	1 (6.7%)	15
W_non_ac_medicine	77 (95.1%)	4 (4.9%)	81
W_ac_medicine	33 (97.1%)	1 (2.9%)	34
W_newsp_other_science	5 (1%)	0	5
W_newsp_brdsh_t_nat_science	7 (1.4%)	0	7
TOTAL	490	70	560

Table 4.3: Frequency of non-agentive VPAs in STEM sections of the BNC

Before closing this section, it is interesting to consider the distribution of non-agentive occurrences across the different sections of the BNC used here. Table 4.3 shows the proportion of agentive and non-agentive uses of VPAs in each of the STEM sections (ordered by the frequency of non-agentive uses). Most non-agentive occurrences are found in natural science and engineering texts, comparatively few in medical texts, and none in newspaper articles. There is no evidence so far of a divide between academic and non-academic writing—instead, stronger discrepancies are observed across scientific domains.

4.5.3 Overall results

Having reviewed the advantages and limitations of both methods, we now turn to the discussion of findings for the entire sample examined, which

4.5. RESULTS: FREQUENCY OF NON-AGENTIVE VPAS

are reported in Table 4.4 below.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Agentive	179 (81.4%)	338 (80.1%)	70 (66.7%)	587 (78.6%)
Non-agent.	41 (18.6%)	84 (19.9%)	35 (33.3%)	160 (21.4%)
TOTAL	220	422	105	747

Table 4.4: Agentive and non-agentive VPAs in the BNC (combined methods)

Unlike in the separate data from methods 1–2 just discussed, a significant effect of agentivity was observed here ($p < 0.05$). *Do that* is significantly more frequent in non-agentive uses than *do it* or *do this* ($p < 0.05$ for both), however *do it* and *do this* have near-identical ratios of non-agentive antecedents and the difference between them is not significant ($p = 0.7$). These data suggest that *do that* is more likely to be used if the antecedent is non-agentive, although this preference was not apparent in the results from each method taken separately.

4.5.4 Influence of finiteness

Table 4.5 below shows the frequency of agentive and non-agentive uses with finite and non-finite VPAs. Contrary to what Houser (2010) and others described for stative *do so*, we do not find a higher frequency of non-agentive VPAs when main verb *do* is non-finite. Instead, although non-finite occurrences are the most frequent, a higher frequency of non-agentive uses is observed in finite cases. The interaction between finiteness and agentivity was found to be significant ($p < 0.001$).

These results mean that the ungrammaticality of non-finite VPE does not affect the distribution of non-agentive antecedents. However, the situation here is quite different from that of stative antecedents, since non-agentive *do it/this/that* are in competition with each other, *do so* as well

as VPE in finite uses. If the antecedent is stative, *do it/this/that* are dispreferred, leaving a choice between *do so* and VPE in finite uses.

	Finite	Non-finite	TOTAL
Agentive	151 (68%)	436 (83%)	587
Non-agentive	70 (32%)	90 (17%)	160
	151	436	587

Table 4.5: Agentivity of finite/non-finite VPAs

[Houser](#) only mentions the influence of finiteness on the frequency of stative *do so*. Although his corpus is also annotated with the agentivity of the antecedent, he does not specify whether non-agentive antecedents are also more likely if *do so* is non-finite. However, this question can easily be answered by examining [Houser's](#) corpus and comparing the frequency of non-agentive cases in finite and non-finite occurrences. The results, reported in [Table 4.6](#), show that non-agentive *do so* is more frequent in finite than non-finite uses (and also more frequent than it is overall), suggesting an effect similar to the one just described for stative antecedents. As with the *do it/this/that* data above, a significant effect of agentivity is observed ($p < 0.05$), meaning that non-agentive *do so* is also affected by the unavailability of VPE in non-finite uses.

	Finite	Non-finite	Total
Agentive	351 (82%)	511 (90.3%)	862 (86.7%)
Non-agentive	77 (18%)	55 (9.7%)	132 (13.3%)
TOTAL	428	566	994

Table 4.6: Agentivity of finite/non-finite *do so* (based on corpus data from [Houser 2010](#))

4.5.5 Influence of register

Contrary to the initial data discussed in Chapter 2, the corpus used here consists mainly of written discourse. Specifically, 679 occurrences of VPAs (90.8% of tokens) come from written texts, and among these 616 are instances of written discourse. However, though spoken data are in the minority, they consistently show higher ratios of non-agentive occurrences. This is true of spoken texts (Table 4.7) as well as spoken discourse generally, including dialogues in written texts (Table 4.8). A significant interaction was observed between agentivity and text type ($p < 0.001$) as well as between agentivity and register ($p < 0.05$).

	Written	Spoken	TOTAL
Agentive	548 (80.7%)	39 (57.4%)	587 (78.6%)
Non-agentive	131 (19.3%)	29 (42.6%)	160 (21.4%)
	679	68	747

Table 4.7: Frequency of non-agentive VPAs in written and spoken texts

	Written	Spoken	Total
Agentive	497 (80.8%)	90 (68.2%)	587 (78.6%)
Non-agentive	118 (19.2%)	42 (31.8%)	160 (21.4%)
	615	132	747

Table 4.8: Frequency of non-agentive VPAs in written and spoken discourse (including written dialogues)

4.5.6 Summary

The data obtained through both methods argue against the claim that VPAs disallow non-agentive antecedents. Although clearly not as frequent as agentive uses, they are attested in a wide range of texts and occur in written as well as spoken speech. In particular, they are encountered in formal

written texts such as academic journal articles or other scientific publications. The spoken sections also include instances of relatively formal register, for instance university lectures, in addition to informal exchanges such as conversations or debates. Non-agentive uses are therefore not as atypical as they might initially appear.

4.5.7 Agentivity and adjuncts

4.5.7.1 Adjuncts

The presence or absence of an adjunct following the VPA, as well as the semantic category of the adjunct, were identified in Chapter 2 as being among the parameters of choice between *do it/this/that*. However, the initial corpus contained mainly agentive occurrences. In what follows we reconsider how the use adjuncts with VPAs is affected by the agentivity of the antecedent. As in the coding of the initial dataset, the annotation scheme only considers adjuncts that directly follow the VPA, and does not take account of the number of adjuncts.

First of all, the data in Table 4.9, showing the frequency of adjuncts with each VPA for the entire sample, are largely similar to those reported earlier in Chapter 2 (see Table 2.5), with a preference for no adjunct overall as well as with *do it/this/that*, and an especially low frequency of adjuncts with *do that*. The interaction between presence of an adjunct and VPA choice is significant ($p < 0.001$). The preference for no adjunct is significant for both *do this* and *do that* ($p < 0.001$), but not *do it* ($p = 0.8$). Unlike in the results obtained with the initial sample, pairwise comparisons between VPAs all turned out to be significant (*do it/do this*: $p < 0.05$; *do it/do that*: $p < 0.001$, *do this/that*: $p < 0.05$).

Data on the frequency of adjuncts across agentive and non-agentive uses (Table 4.10) suggest it is somewhat more common to have an adjunct when the antecedent is non-agentive. The interaction between agentivity

4.5. RESULTS: FREQUENCY OF NON-AGENTIVE VPAS

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Adjunct	108 (49.1%)	170 (40.3%)	27 (25.7%)	305 (40.8%)
No adjunct	112 (50.9%)	252 (59.7%)	78 (74.3%)	442 (59.2%)
TOTAL	220	422	105	747

Table 4.9: Frequency of adjuncts with VPAs, agentivity corpus

and presence of an adjunct is significant ($p < 0.05$), and the preference for no adjunct is significant in agentive uses ($p < 0.001$) but not in non-agentive ones ($p = 0.6$). The above results would appear to corroborate the idea that non-agentive uses of VPAs are not affected by alternation with VPE: having a non-agentive antecedent does not make the use of an adjunct more likely.

	Agentive	Non-agentive	Total
Adjunct	228 (38.8%)	77 (48.1%)	305 (40.8%)
No adjunct	359 (61.2%)	83 (51.9%)	442 (59.2%)
	587	160	747

Table 4.10: Frequency of adjuncts with agentive and non-agentive VPAs

Data for the agentive sample only (Table 4.11) show a preference for no adjunct with *do it/this/that*, which is significant with *do this* and *do that* ($p < 0.001$), but not *do it* ($p = 0.2$). The interaction between presence of an adjunct and VPA choice is significant ($p < 0.05$).

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Adjunct	81 (45.3%)	127 (37.6%)	20 (28.6%)	228 (38.8%)
No adjunct	98 (54.7%)	211 (62.4%)	50 (71.4%)	359 (61.2%)
TOTAL	179	338	70	587

Table 4.11: Frequency of adjuncts with agentive VPAs

In the non-agentive sample (Table 4.12), on the other hand, *do it* shows a preference for having an adjunct, whereas *do that* strongly prefers not

to have one, and *do this* shows no clear preference either way. The results for *do it* and *do that* are significant (*do it*: $p < 0.05$, *do that*: $p < 0.001$). The interaction between presence of an adjunct and VPA choice is also significant ($p < 0.001$). If we compare these data to the overall data in Table 4.9, we see that the ratio of adjuncts is higher with *do it/this* and lower with *do that*. Thus, the presence or absence of an adjunct with *do it* and *do that* respectively appears to be especially important to their non-agentive uses.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Adjunct	27 (65.9%)	43 (51.2%)	7 (20%)	77 (48.1%)
No adjunct	14 (34.1%)	41 (48.8%)	28 (80%)	83 (51.9%)
TOTAL	41	84	35	160

Table 4.12: Frequency of adjuncts with non-agentive VPAs

4.5.7.2 Adjuncts and finiteness with non-agentive VPAs

	Finite	Non-finite	TOTAL
Adjunct	123 (55.7%)	182 (34.6%)	305 (40.8%)
No adjunct	98 (44.3%)	344 (65.4%)	442 (59.2%)
TOTAL	221	526	747

Table 4.13: Frequency of adjuncts with finite and non-finite VPAs, agentivity corpus

The data reviewed in Chapter 2 showed that the frequency of adjuncts varied depending on whether the VPA was finite or non-finite. Specifically, non-finite VPAs are more likely to occur without an adjunct, whereas finite VPAs occur about as often with and without an adjunct (see Table 2.6). The data in table 4.13 replicate this pattern, with a significant interaction between finiteness and VPA choice ($p < 0.001$), and a significant

4.5. RESULTS: FREQUENCY OF NON-AGENTIVE VPAS

preference only in the non-finite/–Adjunct combination ($p < 0.001$), not for finite/+Adjunct ($p = 0.09$).

In the non-agentive sample, however (Table 4.14), we observe a symmetrical effect of finiteness whereby finite VPAs significantly prefer to occur with an adjunct ($p < 0.05$) and non-finite VPAs without an adjunct ($p < 0.05$). This would suggest that the possibility of alternation with VPE in finite uses plays a more important role than when the antecedent is agentive.

	Finite	Non-finite	TOTAL
Adjunct	45 (64.3%)	32 (35.6%)	77 (48.1%)
No adjunct	25 (35.7%)	58 (64.4%)	83 (51.9%)
TOTAL	70	90	160

Table 4.14: Frequency of adjuncts according to finiteness with non-agentive VPAs

Adjunct types

Table 4.15 shows the distribution of adjunct types across non-agentive VPAs. As in the initial corpus (see Chapter 2, Table 2.7), which contained mostly agentive VPAs, we observe the predominance of manner, means and time adjuncts. The most frequent categories with *do it* and *do this* are manner and means adjuncts respectively, as was already the case in the larger corpus. The frequency of these two categories is interesting considering that they may facilitate an agentive interpretation. Purpose adjuncts are not encountered here (17 cases were found in the initial corpus, and 8 with the agentive cases found here). This is explainable by the fact that such adjuncts normally imply volition on the part of the subject and are thus difficult to accommodate with non-agentive VPs.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Manner	11 (40.7%)	12 (27.9%)	1 (14.3%)	24 (31.2%)
Means	3 (11.1%)	17 (39.5%)	0	20 (26%)
Time	5 (18.5%)	5 (11.6%)	2 (28.6%)	12 (15.6%)
Cause	0	5 (11.6%)	1 (14.3%)	6 (7.8%)
Beneficiary	5 (18.5%)	1 (2.3%)	0	6 (7.8%)
Other	3 (11.1%)	3 (7%)	3 (42.8%)	9 (11.7%)
TOTAL	29	42	7	78

Table 4.15: Types of adjuncts with non-agentive VPAs

4.6 Typology of non-agentive VPAs

Having examined the frequency and distribution of non-agentive antecedents across VPAs, we must now turn to the question of what types of non-agentive meanings are attested. The characterisation of agentivity proposed earlier argues that agents must be animates acting intentionally; in other words, they are ‘intentional causes’. The non-agentive category was therefore described as comprising the following four main types: non-volitional animates (animate causes), inanimate causes, and animate or inanimate non-causes. The coding of animacy and causality in the corpus provides a way to determine the frequency of these different types. Results are presented in Section 4.6.1.

It was also suggested that further refinements can be made to the classification of non-agentive occurrences; for instance, non-causal subjects may be (amongst others) patients or recipients, as illustrated earlier in (126)–(129). In order to examine the range of non-agentive situations that may be expressed by VPAs, we also annotated the semantic role of the subject—cause, patient or other—and the specific subtype of situation described, such as natural or material causation, change of state, etc. Section (4.6.2) explains how the semantic role was coded and gives an overview of the distribution of the different roles in the corpus. Sections

(4.6.5)–(4.6.8) provide more detailed discussion of the subtypes distinguished within each category.

4.6.1 Causality and animacy

The data in Table 4.16 show that causal subjects are the most common situation, but non-causal cases still account for over 20% of the non-agentive data. Although *do that* seems more frequent than others in non-causal uses, the interaction between causality and VPA choice was not found to be significant ($p = 0.1$). A significant difference was observed between *do it/do that* ($p < 0.05$), but not *do it/do this* ($p = 0.3$), or *do this/do that* ($p = 0.1$).

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Causal	36 (87.8%)	68 (81%)	24 (68.6%)	128 (80%)
Non-causal	5 (12.2%)	16 (19%)	11 (31.4%)	32 (20%)
Total	41	84	35	160

Table 4.16: Causal and non-causal subjects with non-agentive VPAs

As regards animacy, most non-agentive occurrences in the corpus (84.4%) have an inanimate subject. This is largely a consequence of the types of queries used in method 1, which was primarily intended to exclude occurrences with animate subjects. Data from method 2 (Table 4.17) show a similar but smaller preference for inanimates overall as well as with *do it* and *do this*. However, the interaction between animacy and VPA choice is not significant ($p = 0.7$). The preference for inanimate subjects is significant with *do this* ($p < 0.05$) but not *do it* ($p = 0.1$).

4.6.2 Semantic roles coding

The following semantic roles were used in coding the non-agentive data: Cause, Theme, Patient, Experiencer and Instrument (the larger corpus also

CHAPTER 4. AGENTIVITY

	<i>Do it</i>	<i>Do this</i>	<i>Do this</i>	TOTAL
animate	5 (31.3%)	14 (28%)	2 (50%)	21 (30%)
inanimate	11 (68.8%)	36 (72%)	2 (50%)	49 (70%)
TOTAL	16	50	4	70

Table 4.17: Animacy of VPA subjects (method 2 data, non-agentive occurrences)

contained two occurrences with a Recipient subject, but none were found in the sample studied here). The definition of the Experiencer and Patient roles is relatively straightforward: Experiencer is used with verbs of perception (including mental perception), and Patient applies to entities undergoing a change of state as a result of the event. On the other hand, the use of Cause, Theme and Instrument calls for some clarification.

Cause We assign the role Cause to entities that trigger a particular event or situation, for instance in (134) the presence of clay in water causes a change in the flow of the stream. Causes may be animate (133), in which case they are necessarily unintentional, or inanimate (134), which includes abstract causes (135).

- (133) Mamma is Latin for ‘breast’ and mammals produce milk. Those that evolved in the southern hemisphere, the marsupials, start to do this when their babies are very young indeed. (F9F)
- (134) But deposits of clay can also influence the flow of the stream. They do this inadvertently by changing the level, shape and texture of the ground...(H7X)
- (135) Marriage is one of the elements which leads to the formation and the maintenance of the inequality between classes and it does this at the expense of women. (A6S)

4.6. TYPOLOGY OF NON-AGENTIVE VPAS

Instrument We classify as Instrument any entities used by an animate agent to carry out some action or bring about a certain situation (cf. Fillmore (1968); Schlesinger (1989); these are also called ‘instrumental’ by Dowty (1991) and others). The entity may be a tool in the literal sense, as with the converter in (136), or a method employed to arrive at a desired result (137).

- (136) Existing catalytic converters can produce extremely low hydrocarbon and carbon monoxide outputs only in optimal conditions. The challenge is to develop converters that will do this at once from a cold start and for the life of the vehicle. (FBP)
- (137) One argument has been that the Government’s housing programme was introduced to buy industrial and social peace; it could do this in exchange for really very little, except expense—and when the Treasury, and when the Treasury found it too costly, the scheme was jettisoned. (G05)
- (138) The RITech imaging enhancement firmware certainly does its job well, smoothing out tiny jaggies to give the impression that the print is at a higher resolution than it actually is. Quite how it does it is beyond the scope of this review, but the results are certainly impressive. (HAC)

Theme This label is used for entities undergoing rather causing the action. It includes, amongst others, motion of inanimate entities (139), changes other than changes of state (140), as well as passive exertion of force by the subject, such as in (141) with the animals’ ability to resist cold temperatures.

- (139) There is no doubt that some cells can exhibit chemotaxis, that is, move towards the source of chemical which is diffusing in the medium. Certain white blood cells do this in response to an infective agent. (ASL)

- (140) In the tropics, however, flowers of one kind or another can be found throughout the year, so these animals are able to make pollen and nectar the mainstay of their diet and have evolved highly efficient organs to collect it. Several groups of birds have done this. (F9F)
- (141) Animals also need to cope with the severity of winter, and they have developed a variety of ways of doing this. (FEV)

4.6.3 Results of the semantic role annotation

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Cause	20 (48.8%)	33 (39.3%)	14 (40%)	67 (41.9%)
Instrument	17 (41.5%)	36 (42.9%)	10 (28.6%)	63 (39.4%)
Theme	3 (7.3%)	9 (10.7%)	8 (22.9%)	20 (12.5%)
Experiencer	1 (2.4%)	4 (4.8%)	1 (2.9%)	6 (3.8%)
Patient	0	2 (2.4%)	2 (5.7%)	4 (2.5%)
TOTAL	41	84	35	160

Table 4.18: Semantic roles with non-agentive VPAs

Results of the semantic roles annotation appear in Table 4.18. Instrument, Cause and Theme are the most frequent roles for non-agentive VPA subjects; Experiencer and Patient are attested only rarely. The predominance of the Cause and Instrument roles (the latter being a specific subtype of cause) suggests, along with the data in 4.6.1, that the subject of non-agentive VPAs is most often an inanimate cause. This is unsurprising given the semantics of main verb *do*—an entity which causes an event is more likely to be viewed as doing something than one which is affected by the event. The existence of cases with Theme or even Patient and Experiencer subjects nevertheless suggests that certain non-causal situations can still be conceived of as instances of ‘doing’. Lastly, four occurrences

of *do this/that* can be regarded as having a stative antecedent. These are discussed further in (4.6.9).

The following sections will explore the types of non-agentive situations expressed by VPAs in further detail. Due to the small size of the sample (especially for *do it/that*) we only provide an overview of the overall frequency of the subtypes identified for each semantic role, and do not go into the details of variation between VPAs in this respect.

4.6.4 Causes

Non-volitional animate causes are grouped under the label ‘unintentional animates’. Inanimate causes, as suggested above, can be further subdivided into first-order entities (143) and abstract ones, such as an action or event (144), or a process (145), or a state (146). We annotate the former as ‘material causes’, and the latter as ‘eventive’ and ‘stative’ causes respectively. We classify entities that are causally involved in naturally-occurring processes, for instance (143), as ‘natural agents’. The term is mentioned by Cruse (1973, p. 11) as a typical way of referring to forces of nature, primarily those linked to weather events: “[W]e commonly describe the sun, wind, frost, etc., as ‘natural agents’ [...]”. Similarly, Lowder and Gordon (2015) describe as ‘natural forces’ those inanimate entities that are capable of generating their own energy. Our usage of this category is somewhat broader and also covers, for instance, elements triggering a chemical reaction as is the case in (143).

- (142) Because the glass in a greenhouse traps the sun’s energy (though it does this mainly by inhibiting convection, thereby stopping warm air rising and escaping), this process has come to be known as the ‘greenhouse effect’. (GU5)
- (143) ECT increases the central nervous system’s turnover of norepinephrine, and does it more quickly with fewer side-effects than the

CHAPTER 4. AGENTIVITY

tricyclic drugs. (B7K)

- (144) The emergence of effective multimedia capabilities will provide a new dimension of utility for electronic information products and may hasten the growth and development of their markets. However, it will not do this simply because people forecast that it will. (HRD)
- (145) There is no subtler, no surer way of overturning the existing basis of society than to debauch the currency. The process engages all the hidden forces of economic law on the side of destruction and it does it in a manner which not one man in a million is able to diagnose. Already we can see this process at work. (CWD)
- (146) Marriage is one of the elements which leads to the formation and the maintenance of the inequality between classes and it does this at the expense of women. (A6S)

SUBTYPE	
Natural agent	24 (35.8%)
Material cause	15 (22.4%)
Stative cause	10 (14.9%)
Unintentional animate	8 (11.9%)
Eventive cause	7 (10.4%)
State	3 (4.5%)
Total	67

Table 4.19: Types of causal subjects with non-agentive VPAs

Table 4.19 shows the frequency of different types of causes. It appears that the subject is in most cases a natural agent. The fact that such entities act using their own energy may lead them to be perceived as animate and possibly even intentional in much the same way as true animate agents, as [Lowder and Gordon \(2015\)](#) remark:

4.6. TYPOLOGY OF NON-AGENTIVE VPAS

Natural forces are semantically inanimate (nonliving), but behave in ways that are more similar to animates than inanimates in that they are able to initiate movement, change course without warning, and occasionally cause destruction, injury, and death.

Our data suggest the similarity extends to other kinds of natural agents we have distinguished. This is apparent in the use of verbs such as *exert* and *influence* in (147), or the occurrence of *do this* in a purpose adjunct in (148).

(147) DNA patterns, or genes, exert their effects by influencing the course of events in the chemical factory, and they do this via their influence on the three-dimensional shape of protein molecules. (H7X)

(148) Each kind of protein machine churns out its own particular chemical product. To do this it uses raw materials that are drifting around in the cell, being, very probably, the products of other protein machines. (H7X)

Material entities (with the exception of machines, discussed in the next section), states and events typically do not exhibit the same proximity to animates, and there is usually nothing in the context suggesting that the subject possesses some degree of apparent animacy or intentionality (149).

(149) When the smoke really thins out, thin blue smoke, rap on it, give the whole tin a shake and it'll suddenly burst into thick smoke for a few minutes and it may do that again for a second time. (EHF)

However, certain events or situations may involve the action of animate agents, such as debauching the currency in (145), so that the process described by the VPA is ultimately caused by an intentional action. This reasoning is also applicable to certain unintentional actions carried out by animates, such as the outcome of a dice throw in (150)

CHAPTER 4. AGENTIVITY

(150) Near this left-hand end of the scale are things that are only slightly improbable, such as shaking a double six in a single throw of a pair of dice. The odds of this happening are 1 in 36. I expect we've all done it quite often. (H7X)

A less obvious instance of causation is exemplified in (151). While the sentence as a whole ultimately describes the contents of the books in question, the antecedent-trigger *tells a story* emphasises the act of narration. Though this may be regarded as a case of metonymy for the author's intention (strictly speaking it is the author, not the book, who is telling a story), the choice of the inanimate subject *every page* suggests that what is described is the effect of the book on its readers; the antecedent may be paraphrased as something like 'convey a story or message'.

(151) The approach of all these Usborne books is to introduce the reader to the fun that can be had with electronics. The presentation is crisp and the topics broken down into easily comprehensible parts: every page tells a story and does it in such a delightful way that the reader is led on through the book. (B7H)

4.6.5 Instruments

SUBTYPE	
Pseudo-agent	47 (74.6%)
Tool	16 (25.4%)
TOTAL	63

Table 4.20: Types of Instrument subjects with VPAs

Most occurrences where the subject has the role of Instrument involve a special class of inanimates which we call 'pseudo-agents', a cover term for various types of machines or devices that are to some degree autonomous, capable of acting independently of human control. Like natural agents,

4.6. TYPOLOGY OF NON-AGENTIVE VPAS

they are in fact near-animate, but they are even more likely to be perceived as animate and intentional to some degree, especially when performing tasks that would otherwise be carried out by human agents. The most common examples of pseudo-agents in the corpus are computer systems and programmes (152), but we also find occurrences with various kinds of appliances such as a microwave (153) or a VCR (154).

- (152) If you are scoring the tests yourself, rather than getting the computer to do it for you, then always do so at the same time of day—otherwise there would be variability due to time-of-day effects upon your scoring ability (A75)
- (153) [I]f you were doing a jacket top potato [with a microwave], you er would weigh the potato and, and then erm put in the weight, say it was six ounces, and all you do then is press erm jacket potato,(pause) jacket potato, you don't have to put any time, it automatically does it for the time. (KBC)
- (154) The VCR can play a tape of one standard, either NTSC or PAL, display it on any PAL or NSTC TV set, and make a copy onto any other VCR in either PAL or NTSC. The only other VCR on the market which can do this is made by Panasonic and costs £1800.. (C92)

The following examples further illustrate how pseudo-agents are often described as behaving in a way that mirrors the actions of human agents; for instance making attempts (155) or decisions (156).

- (155) If you are using a full 24 bit colour photo-retouching or vector graphics drawing program on a system which can only display 16 or 256 colours Windows will attempt to represent all the screen colours required. It does this by a process called dithering which takes two or more colours from the 16 or 256 colour palette available and places them side by side to give the impression of another

colour. (HAC)

- (156) Of course, it was necessary to come to a decision at some point, and to do this the system relied primarily on the acoustic probability scores attached to each labelling. (HX9)

Also classified as Instruments are non-autonomous entities used as tools. These may be tools in a concrete sense, such as a drill (157), or methods employed to achieve a particular goal, for instance a communication policy (158).

- (157) The drill is basically designed for drilling holes into extremely hard masonry, usually for some type of fixing, or to clear a way for cables or pipes. It does this extremely efficiently, but if you want to use the tool for conventional drilling and screwdriving, the hammer action must be switched off [...]. (A16)
- (158) For example, the purpose of this company's communication policy is relatively simple. It is to help us reach our organization's main objectives. It does this through encouraging a favourable internal climate of opinion in which all employees can feel involved with, and understand, the firm's affairs. (AYJ)

4.6.6 Themes

As shown in Table 4.21, the majority of Theme subjects are inanimate entities undergoing a change of some kind, most often a change of location (159). Two exophoric examples of *do that* are attested in this category (160)—(161).

- (159) A compass is really a magnetized needle which always points to the North. It does this because there is a large area of magnetized rock near to the North Pole which attracts the needle. (G25)

4.6. TYPOLOGY OF NON-AGENTIVE VPAS

SUBTYPE	
Motion	12 (60%)
Change	4 (20%)
Passive ability	3 (15%)
State	1 (5%)
TOTAL	
	20

Table 4.21: Types of processes denoted by VPAs with Theme subjects

- (160) ‘But why did the keys fall faster than the paper?’
 ‘Because they’re heavier, of course.’
 ‘Ye-es,’ said Uncle Albert. ‘They’re certainly that. But there’s something else going on here. Take another look. Keep your eye on the paper.’ As it fluttered down again he said, ‘There! Why is it doing that?’
 ‘Going from side to side you mean?’
 ‘Yes.’ (FNW)
- (161) He opened a bin that said EMERGENCY in big red letters. A large dispenser of detergent, an old flex microphone as big as a taper, a polythene bag of tubular bandages, a sealant gun and a box of raisins fell all over his feet. ‘It always does that,’ said Tabitha. (CJA)

Three of the four examples describing other types of changes have to do with the evolution of animal species. Although the process is not as such controlled by the subject, the context may sometimes give the impression that it is, as with the adjuncts *most successfully* in (163) or *in a different way* (164). In the latter case, moreover, the phrase *perfected a different gait* suggests effort and practice to achieve the desired result, as if the animal had selected this particular way of moving.

- (162) In the tropics, however, flowers of one kind or another can be found throughout the year, so these animals are able to make

CHAPTER 4. AGENTIVITY

pollen and nectar the mainstay of their diet and have evolved highly efficient organs to collect it. Several groups of birds have done this. (F9F)

(163) However, this is not to say that many forms of life have not been able to adapt to life on slopes up to the 10,000-ft (3,050 m) level, and in some cases higher. One of the true leopards has done this most successfully. The snow leopard, *Panthera uncia*, has adapted to life among the screes and snows at altitudes of up to 20,000 ft (6,100 m). (CK2)

(164) Grassland life is much the same the world over, and horses and litopterns independently evolved the same qualities to cope with the problems of grassland life. In particular, the litopterns, like the horses, lost all their toes except the middle one on each leg, which became enlarged as the bottom joint of the leg and developed a hoof. The leg of a litoptern is all but indistinguishable from the leg of a horse, yet the two animals are only distantly related. In Australia the large grazers and browsers are very different—kangaroos. Kangaroos have the same need to move rapidly, but they have done it in a different way. Instead of developing four-legged galloping to the high pitch of perfection that horses (and presumably litopterns) did, kangaroos have perfected a different gait: two-legged hopping with a large balancing tail. (J52)

(165) shows a different context; it may be analysed as a case of metaphorical motion accompanied by a change in status of the phrase *I love you*.

(165) ‘I love you’ shouldn’t go out into the world, become a currency, a traded share, make profits for us. It will do that if we let it. (G1X)

4.6.7 Experiencers

Six occurrences in the corpus describe non-agentive cognition in the sense that the experiencer does not make a conscious effort to perceive a stimulus. We found no occurrences describing perception through any of the five senses; the antecedent-trigger contains either a mental verb such as *know* (166) or *identify* (167), or a sensory perception verb that is interpreted as purely cognitive (see in (168) is to be understood as ‘recognise’ rather than ‘visualise’). It is arguable that some cases of mental processing, particularly ‘identification’ contexts like (167)–(168) may be seen as requiring a certain effort on the part of the subject, however this is clearly not applicable in (166).

- (166) In this chapter we are looking a little more closely at speech, the ‘twin’ of speech reading. You already know and experience speech subconsciously because you have been doing it successfully for years and years and years. (C9R)
- (167) Similarly, we can identify sounds with our eyes closed, even distinguishing between those which are very alike. The brain does this by recognising the rhythm and pattern of the sounds, including their frequencies. (FEV)
- (168) The final multifaceted perfection of mimicry has been put together by the summed natural selection provided by many different species of predators. No one predator sees the whole perfection of mimicry, only we do that. This seems to imply that only we are “clever” enough to see the mimicry in all its glory. (J52)

4.6.8 Patients

As noted in (4.6.3), Patient subjects in the corpus are only encountered with *do this/that*; however, one example with two subsequent occurrences *do it* (171) was found online.

- (169) We might therefore suppose that it is normal for the growing bough to straighten under stress by some simple non-living mechanism. However, not all plants do this [...] (CEG)
- (170) At some time or another most of us will have sat beside the shore of a loch when there's complete calm and not a breath of air. And you know, you look at the water and you see in the water a perfect reflection of what there is above. And then, almost imperceptibly, the calm surface becomes ruffled and the clear images blurred, and it becomes colder. And it does that because a wind has sprung up, you can't see it but you can observe its effect. (G5H)
- (171) Even if the ice doesn't completely disappear, it is very likely that this will be a record low year. I'm convinced it will be less than 3.4 million square kilometres. I think there's a reasonable chance it could get down to a million this year and if it doesn't do it this year, it will do it next year. (*The Independent*, June 4, 2016)

4.6.9 Stative antecedents

Four occurrences of *do this/that* in the corpus have a stative antecedent; they appear below in (172)–(175). No occurrences of stative *do it* were encountered, and it does not seem to be acceptable in any of the examples below. The stativity of the antecedent is shown by the use of the simple present and the impossibility of the progressive, e.g., **The inference is preserving the truth of its premise*. Contrary to the other non-agentive uses examined so far, stative antecedents with *do it/this/that* have consistently been judged impossible in the literature, and no attested examples have previously been reported.

- (172) For an indirect observation to be good, therefore, both parts of it must be good. The direct observation must be good, and so must the inference. That is, the inference must preserve the truth of

4.6. TYPOLOGY OF NON-AGENTIVE VPAS

its premise (that the sign is present) in its conclusion (that what the sign signifies is present). And it will do that just in case, as a matter of fact, the sign is correlated [...] (FBD)

- (173) The sceptical argument therefore claims that you cannot make sense of the idea of a subject of experience other than yourself. [...] Why does this show that the argument from analogy, by making assumption (a), cannot account for (b), our understanding of statements about others? It does this because it shows how if we start from our own case alone and concentrate entirely upon a conception of mental states which is independent of behaviour, we can not move from our conception of ourselves as subjects of experience to a conception of other subjects. (F9K)
- (174) An important point to note is that this simplification of semantics is not just a reduction of problems in the lexicon; it also makes possible the adoption of a semantics built on simple logical principles. It does this by demonstrating that once pragmatic implications of the sort we shall call implicature are taken into account, the apparently radical differences between logic and natural language seem to fade away. (J2K)
- (175) In other words, multimedia must confer real benefits. It must suit the purpose for which it is used. If it fails to do this, multimedia will be an irrelevance. (HRD)

A common feature of the first three examples is that the subject is a cause, specifically one that does not involve action. This is apparent with *makes possible* in (174), but *preserve* in (172) can be interpreted as ‘cause to remain’, and likewise *show* in (173) means ‘make clear’. Since previous sections have shown that most non-agentive uses involve causation, it is likely that this feature is essential to the acceptability of the examples. The case of (175) is more complex, however: while *confer benefits* clearly

involves causation, *suit our purposes* implies a judgement on the adequacy of the material.

Two of the stative cases are non-finite (175–172), and three have a non-contrastive adjunct (172–173). Since there is no finite adjunct-less case, VPE with *do* is excluded, however it may occur after *to* in (175) (*If it fails to*). *Do so*, which allows stative antecedents in certain cases, appears to be acceptable in all of the above.

4.6.10 A note non-on on agentive *do so*

As discussed in (3.1.2), Quirk et al. (1985) argue that *do it/this/that* can have non-agentive antecedents, but *do so* will systematically be more acceptable in such cases. Although Houser (2010)'s data suggest that *do so* is more frequent in non-agentive uses (13.3% of 994 occurrences, as opposed to just about 2% in our 900-occurrence corpus), it is not necessarily the preferred form. As shown by Miller (2013), it is subject to independent discourse constraints and becomes less acceptable if these are not respected. These constraints apply irrespective of the agentive or non-agentive status of the antecedent. The following examples suggest that *do so* is dispreferred with beneficiary adjuncts (176), remnant complements (177) or in certain adjunct-less contexts (178).

- (176) If you are scoring the tests yourself, rather than getting the computer to do it/#so for you [...] (A757)
- (177) All the time I would like to be somewhere else, someone else. But I'm not. I'm me. I can't escape it. It contaminates my sleep. It destroys my digestive system. How has this happened? Why is it doing this#so to me? (A08)
- (178) 'I've decided on a different plan, one that will keep us safely out of temptation's way.' There was only one plan that would do that/#so, Caroline thought, and it was hers. (JY7)

In many others, such as (179)–(181), *do so* is felicitous but not particularly more acceptable than *do it/this/that*. Note moreover than in (179), *do so* is used in the agentive reading of the antecedent *re-engineer applications* (with *commercial users* as the subject), and *do it* in the non-agentive reading.

- (179) Until now, ICL says, commercial users have had to re-engineer applications to take advantage of parallel CPU architectures, and have proved unwilling to do so. ‘The moment database vendors come out with versions that do it/so automatically, the rules change,’ it says. (CSJ)
- (180) These so-called superantigens have been named for their ability to activate simultaneously large numbers of T cells. Superantigens *do this/so* by bypassing the normal route of intracellular processing and binding directly as intact proteins to class II MHC molecules [...] (HWW)
- (181) The creation of the CBO began to redress the balance of power. It *did that/so* via one fundamental way—it ended the president’s monopoly on information, on budget forecasts, on economic forecasts. (EAY)

4.6.11 Acceptability of non-agentive VPAs

The findings presented in the previous sections show that there are cases where non-agentive VPAs are not markedly less acceptable than agentive ones. However, in most attested cases the subject is at least the cause of the action, and main verb *do* can express simple causation independently of its use in VPAs. Moreover, two common types of inanimate subjects in the corpus, natural agents and pseudo-agents, tend to behave in ways similar to animate agents, a property which is often reflected in the linguistic description of such entities. While causality and similarity with

animates are two important factors in licensing non-agentive uses, their acceptability is also linked to independent factors favouring the use of a particular VPA identified in Chapter 2, such as the presence of a non-contrastive adjunct—especially manner and means adjuncts with *do it* and *do this* respectively—or the absence of an adjunct following *do that*.

Although a few of the examples in the spoken sections clearly belong to the informal register (183 also being an example of non-standard usage, as shown by %*I've been paining me*), there is no reason to treat them as performance errors; rather, they may be said to illustrate the typical usage of VPAs, aside from the antecedent being non-agentive. (182) shows a typical pattern where *do it* is followed by *again*, whereas (184) has *do that* in an adjunct-less *if*-clause. In (183), *do it* occurs in a cleft sentence with a contrast between two possible subjects:

- (182) — What about your leak that came through the ceiling, is that cleared now?
— Oh it's doing it again it's doing it again, dear (pause) oh yes, doing it (KBF)
- (183) — And I've been up till, I've been paining me all night Doctor, Aye. I'm just wondering if it's my teeth or that it's just my blood that's doing it.
— I think I think it might be s– the teeth (FXN)
- (184) Now the danger of leaving these is you can get– i– is the infection can spread into the bone soon. And if it does that you're in real trouble. (G5N)

Similarly, VPA choice in non-agentive contexts often seems driven by parameters unrelated to the agentivity of the antecedent. For instance, *do that* is preferred over *do it* or *do this* in polar questions, which explains why it occurs in (185), a simplified variant of (125). *Do it* becomes felicitous in a manner interrogative (186) or if followed by a manner adjunct (187).

- (185) The purpose of this machine is to save stress on the individual.
— And does it actually do that/#it?
- (186) And how does it do it?
- (187) The purpose of this machine is to save stress on the individual, and it does it/this/that quite efficiently.

As regards the difference in acceptability between different types of non-agentive uses, it is likely that causal ones, which are consistent with the semantics of *do*, will be preferred over non-causal ones. However, even non-causal antecedents can become acceptable if the context favours the use of particular VPA for independent reasons.

4.7 Conclusion

The data examined in this chapter allow us to reject the hypothesis that *do it/this/that* require an agentive antecedent, as proposed by [Huddleston and Pullum \(2002\)](#) and [Culicover and Jackendoff \(2005\)](#) amongst others. Although the subject is most often an agent in the narrow sense of an intentional animate entity, it is sufficient that it be a cause of the antecedent action, either animate or (in the most common case) inanimate. As argued, this is in fact predictable given the semantics of main verb *do*, which is itself not restricted to agentive situations. Despite the frequency of causal cases, however, it would be incorrect to claim that the antecedent must be causal, rather than agentive, as non-causal uses are also attested, with subjects having such semantic roles as Theme, Patient, Experiencer or Recipient.

VPAs can be used to express a wide range of non-agentive situations, most often involving actions caused by unintentional animates, or inanimates. In the latter case, the subject often comes close to an animate agent insofar as it possesses (or is at least perceived as possessing) some

degree of autonomy and control over its actions. This resemblance to actual agentive cases has been argued to be essential to the acceptability of such examples. More generally, it has been shown in 4.6.11 that non-agentive VPAs are licensed by factors unrelated to the agentivity of the antecedent, which also apply to agentive uses. In particular, several of the non-causal examples have contexts that are otherwise typical for the VPA in question, which would explain why they sound more natural than the constructed examples given by Huddleston and Pullum or Culicover and Jackendoff.

As regards VPA choice in non-agentive contexts, *do that* has been shown to occur with a non-agentive antecedent more often than *do it* or *do this*, which do not differ significantly from each other in this respect. Further research is needed to understand this difference and determine its importance in the choice between VPAs. It has also been argued that, contrary to the claims of Quirk et al. (1985), *do so* is not always more acceptable than *do it/this/that* when the antecedent is not agentive—this will only be the case if the relevant discourse conditions for its use are met, otherwise *do it/this/that* are preferred. More research is needed on the alternation between *do it/this/that*, *do so* as well as VPE in non-agentive contexts.

Finally, our data also revealed the existence of stative uses, which had previously been deemed impossible. Though they can be regarded as exceptional, the attested examples collected here are nevertheless felicitous, as in all but one the subject is a cause, hence of the typical category for non-agentive VPAs. The exact conditions on stative antecedents and their acceptability would need to be established based on a larger sample than the four occurrences in our corpus. The specific queries used to build the corpus used here proved efficient in retrieving non-agentive occurrences, but they are unfortunately inadequate to search for examples of stative antecedents, since the stative or non-stative aspect of the antecedent is unaffected by the animacy of the subject. However, there seems to be no

4.7. CONCLUSION

obvious alternative method that would facilitate a search for stative examples. It is also uncertain that many more such examples could be found, considering that the larger corpus does not contain any.

Chapter 5

Antecedent mismatch

5.1 Introduction: Mismatch with VPA and VPE

This chapter is concerned with ‘antecedent mismatch’ under VP anaphora, i.e., when the structure of the antecedent-trigger is not parallel to that of the VPA clause. This may be due to a voice mismatch—for instance, a passive antecedent-trigger with an active VPA (188)—or a category mismatch if the trigger is not a VP (as in (189) where it is an NP).

(188) It was clear that funds were being embezzled but who could be doing it? Could it possibly be the accountant Edward Morris?
(CKD)

(189) In essence Groupware requires customisation since business processes differ office to office, industry to industry. Uniplex is putting itself in a position to do that internally with the acquisition three weeks ago of IMI Consulting [...] (CNJ)

Alternatively, the antecedent may be inferred from the discourse context (190) or, in exophoric uses, directly from the speech situation (191). Such examples may be described as having an ‘inferred antecedent’. In the first of these (190), *do it* means not just ‘retire to various parts of the

5.1. INTRODUCTION: MISMATCH WITH VPA AND VPE

country' but specifically to the more pleasant areas. In (191), *do this* is interpreted as referring to the activity which the speaker is currently engaged in, i.e. cooking for the others.

- (190) They're retiring to various parts of the country, down here onto the Sussex coast and on the whole coast into the South West—to the pleasant places if they can afford to do it [...] (KRE)
- (191) Rodney cracked two eggs into the frying pan. 'I'm not doing this every morning,' he said. 'With so many of us we ought to set a rota for cooking.' (AOR)

Mismatches such as the above have been the subject of much discussion in the literature on ellipsis, which has traditionally been thought to require a an identical antecedent at the relevant level of structure, but comparatively little has been said about mismatch with VPA since they are usually considered exempt from this constraint. Though a number of studies have commented on the effects of mismatch on VPE and *do it*, none have addressed the specifics of how different VPAs behave in this respect. Also, whilst passive/active mismatches are a common topic, fewer studies have examined exophoric and other 'antecedentless' uses of VPAs, as they have traditionally been described.

Much of the debate around antecedent mismatch with VPE has been informed by the observation from [Hankamer and Sag \(1976\)](#) that certain anaphors, such as VPE, appear to require a syntactically parallel antecedent-trigger, while others, such as *do it*, do not. This conclusion was made on the basis of examples such as (192), where active VPE is much less acceptable than *do it* in the context of a passive antecedent-trigger:

- (192) The oats had to be taken down to the bin,
a. *so Bill did.
b. so Bill did it (65a-b, p. 413)

On the basis of this and other (invented) examples, [Hankamer and Sag \(1976\)](#) proposed that anaphors could be divided into two classes—‘deep’ and ‘surface’ anaphors—according to whether or not they allowed antecedent-trigger mismatch or exophoric uses with no trigger present. This central claim prompted a number of studies on antecedent mismatch in the psycholinguistics literature, dealing with VPE only (e.g., [Arregui et al. \(2006\)](#); [Kertz \(2010, 2013\)](#); [Phillips and Parker \(2014\)](#)) or VPE and *do it* and sometimes other types of anaphors such as NCA ([Murphy, 1985, 1990](#); [Tanenhaus and Carlson, 1990](#); [Maurer et al., 1995](#); [Roberts et al., 2013](#)). The shared aim of all these papers was to test empirically the claims of [Hankamer and Sag](#) and specifically the prediction that deep and surface anaphors are processed differently. Interest in this issue increased with the discovery of attested cases of mismatch with VPE, for instance the passive/active mismatch in (193) reported by [Hardt \(1993\)](#). This led various authors to consider how such examples differ from constructed ones like (192).

(193) This information could have been released by Gorbachev, but he chose not to. (Daniel Schorr, NPR 10/17/92) [ex 131, p. 37]

Attested examples with VPA have also been recorded (at least with *do it*) though understandably they have attracted little attention since they are perfectly consistent with the predictions of [Hankamer and Sag](#). [Souesme \(1985\)](#) provides the following example of *do it* with a nominal antecedent-trigger.

(194) The rapid emergence of the IBM standard was far more than anticipated. IBM did it with an awesome speed, which took most people unawares. (114, p. 50)

In this chapter we will discuss further such examples based on corpus data from the BNC, and how these compare to constructed ones used in experimental studies. The rest of the chapter is organised as follows: Sec-

tion 5.1.1 explains the deep/surface anaphora distinction from [Hankamer and Sag](#) in greater detail. Section (5.1.2) reviews a series of psycholinguistic accounts of mismatch, primarily those which address VPE and *do it* jointly. Section (5.2) presents the findings of a corpus study of mismatch with VPAs focusing on passive and nominal antecedent-triggers. Inferred antecedents, including exophoric uses, are discussed in Section 5.3.

5.1.1 Deep and surface anaphora

In a seminal paper, [Hankamer and Sag \(1976\)](#) point out that while certain kinds of anaphoric expressions can be felicitously used without a linguistic antecedent (an antecedent-trigger, following [Cornish \(1999\)](#)'s terminology which we use here), others must occur with a linguistic antecedent in all cases. Expressions of the first kind, which they describe as 'pragmatically controlled', include third-person pronouns as well as VPAs such as *do it*. The other type, of which VPE is an instance, is called 'syntactically controlled'. They illustrate this distinction with the following pair, in which VPE is infelicitous in the absence of an antecedent-trigger, whereas *do it* is acceptable:

(195) [Hankamer attempts to stuff a 9-inch ball through a 6-inch hoop]
Sag: #It's not clear that you'll be able to. (ex. 3, p. 392)

(196) [Same context]
Sag: It's not clear that you'll be able to do it. (ex. 4, p. 392)

Linguistic mention of the antecedent, as in (197), improves acceptability with VPE.

(197) Hankamer: I'm going to stuff this ball through this hoop. Sag: It's not clear that you'll be able to. (ex. 5, p. 392)

Despite the badness of (195), [Hankamer and Sag](#) do not regard it as ungrammatical, as signalled by the cross-hatch symbol:

[T]he anaphoric process that leaves a pro-form *do it* as anaphoric VP can more readily be pragmatically controlled than the process known as VP Deletion, which leaves behind no pro-VP, but only a bare Aux or stranded complementizer. (p. 392)

They also observe that VPE appears to require an antecedent that is syntactically parallel to the VPE sentence. Contrary to the exophoric uses above, however, they consider mismatches with VPE such as seen earlier in (192) to be ungrammatical.

This and other evidence led [Hankamer and Sag](#) to propose that there are two classes of anaphors: deep anaphors, in which the anaphor is present in deep structure, and surface anaphora, in which it is derived through deletion or pronominalisation, and not present in deep structure. Specifically, the distinction accounts for whether syntactic identity is required for the interpretation of anaphoric expressions: surface anaphors requires that the antecedent and anaphor be syntactically identical, whereas deep anaphors requires no identity at all, only that the antecedent be recoverable from the extended context. In a later paper, [Sag and Hankamer \(1984\)](#) elaborate on these proposals and make specific hypotheses on the processing of both anaphor types. The revised distinction is as follows (p. 328) :

The interpretation of a ‘deep’ anaphoric element is determined by reference to the interpretation of its antecedent (in all cases when there is one), i.e, by reference to some object in a model of the world constructed by the interpreter of the sentence of discourse; while the interpretation of a surface anaphoric element is determined by reference to a linguistic representation associated with the antecedent, specifically a propositional representation of the kind generally called logical form.

It follows from this that surface anaphors are expected to incur a penalty for antecedent mismatch, since the syntactic identity condition does not

obtain, whereas deep anaphors are supposed to be unaffected by mismatch. However, several authors such as [Hardt \(1993\)](#) and even [Sag \(1976\)](#) himself reported attested cases of mismatch with VPE. An example of a passive/active mismatch was mentioned in (193) earlier; below is a further one from, also from [Hardt](#), showing a nominal antecedent-trigger:

(198) We should suggest to her that she officially appoint us as a committee and invite faculty participation. They *won't*, of course,...
(UPENN email message) (ex. 116, p. 35)

The status of such cases has been a matter of debate in the literature. Some authors take these as evidence that VPE does not in fact require a syntactically parallel antecedent, while others consider them as ungrammatical but 'repaired' by the addressee ([Arregui et al., 2006](#)), inducing a higher processing cost. Others still argue that mismatch is grammatical but its acceptability is subject to general constraints on parallelism as well as specific discourse conditions on the use of ellipsis ([Kehler, 2000, 2002](#); [Kertz, 2010, 2013](#); [Miller and Hemforth, 2014](#)).

By contrast, mismatch with VPA is a much less controversial issue, since it is predicted to be always acceptable. For this reason, there have been no detailed analyses of the behaviour of VPAs with non-parallel antecedents, and although *do it* is commonly used in experimental studies on ellipsis as a prototypical deep anaphor and taken as representative of the entire set of *do NP* VPAs; possible differences with *do this/that* are ignored. Since all are considered as deep anaphors, it is implicitly assumed that they will behave identically with respect to mismatch, making an experimental analysis of mismatched VPAs unwarranted.

5.1.2 Experimental accounts of mismatch with VPE and *do it*

[Murphy \(1985, 1990\)](#) was among the first to investigate the deep/surface

distinction experimentally, taking *do it* and VPE as representative of each type. [Murphy \(1985\)](#) examined the processing of VPE and *do it* depending on the length, distance and parallel or non-parallel structure of the antecedent-trigger (what he calls ‘syntactically inconsistent antecedents’). The hypothesis was that longer, more distant or non-parallel antecedent-triggers should all induce a higher processing cost (as measured by reading times) with VPE but not with *do it*, on account of the need for copying of linguistic structure with the former but not the latter. However, [Murphy](#) showed that distance and length of the antecedent-trigger also slowed down the processing of VPA, not just VPE, and further that there was no reliable difference in the processing of both constructions under mismatch. Thus the findings bring partial support for [Sag and Hankamer](#)’s hypothesis, insofar as VPE is affected by mismatch, however the fact that *do it* is just as sensitive to mismatch is unexpected.

[Murphy \(1990\)](#) reconsidered the effects of distance in an experiment involving a reading task and a sensicality judgement task. The effect of distance was tested by the presence or absence of a ‘filler’ sentence between the sentence with the antecedent and the sentence with the anaphor. A further variable that was explored depended on whether or not the intervening filler sentence provides a competing antecedent for the anaphor, thereby making the target sentence ambiguous. Since the antecedent in the ambiguous condition is stative, it is inappropriate for *do it*, meaning that ambiguity may arise only with VPE. [Murphy](#)’s materials constructed according to this design were embedded inside larger 8-line stories. An example text is reproduced below (p. 681-2 of the paper)

Ellen Marcovitz was flying from Seattle to Washington DC.
She was a sales representative for Acme Aviation, Inc.,
and she was trying to get a big government contract for her company.
Because she was nervous, she started a conversation with the man

5.1. INTRODUCTION: MISMATCH WITH VPA AND VPE

next to her.

The man asked her a question about Acme Aviation. [antecedent]

— This was more relaxing than her previous worrying. [unambiguous filler]

— She felt more relaxed almost immediately. [ambiguous filler]

Later, Ellen wondered why he did/did it. [anaphor]

She worried that she might have told something important to this stranger.

Results for the reading experiment showed faster processing of close than distant antecedents, but no difference between VPE and *do it* and no effect of ambiguity. Murphy noted that the ambiguous conditions actually tended to be read faster, an effect which he explains, following [Garnham and Oakhill \(1987\)](#), by suggesting that subjects select the competing antecedent in the intervening sentence even if it is implausible. In a second experiment, [Murphy](#) tested the same set of materials using a task in which participants are asked to judge whether the test sentences make sense in the context of the preceding story. [Murphy](#) does not specify what ‘make sense’ means for the purposes of this task, but it seems a poor indicator of acceptability or grammaticality. First, it is well known that ungrammatical sentences can make sense, and secondly, participants may well attempt to make sense even of implausible sentences if they are invited to do so. Murphy found that anaphors were judged to make sense more often when the antecedent was near, but there was little effect of ambiguity. He observed no effect of anaphor type on sensicality judgements, though it affected judgement latencies, with an increase in reaction times from the close to the distant-ambiguous condition with VPE, and slightly shorter reaction times with *do it*.

In short, while results from the second experiment bring some support for [Sag and Hankamer](#)’s proposals, those from the first experiment do not. [Murphy](#) suggests this is a consequence of the nature of the tasks used, with subjects paying greater attention to grammaticality when judging if

a sentence makes sense as opposed to simply reading it for comprehension.

Tanenhaus and Carlson (1990) criticised the methods used by Murphy, claiming in particular that length variations can introduce scope and attachment ambiguities which may independently account for the reported length effects. Specifically, they argue that the intervening sentence may remove the antecedent from focus. According to them, if the antecedent is kept in focus, there should be no effect of distance. Tanenhaus and Carlson tested passive/active mismatches as well as nominal antecedents with VPE and *do it* using a ‘make sense’ task similar to Murphy’s design above. In a third experiment, they also compared VPE and null complement anaphora (NCA), which although elliptical is considered a deep anaphor (see Hankamer and Sag (1976) for the reasons why; amongst others, it easily allows passive/active mismatches e.g., *The rubbish has to be taken out, but Bill refused*, as well as exophoric uses: *Try harder!*).

Results for the first two experiments were very similar, and Tanenhaus and Carlson observed an effect of anaphor type and parallelism: surface anaphors, but not deep anaphors, were judged to make sense less often, and with longer judgement latencies, when the antecedent was non-parallel. This being said, the proportion of sentences judged to make sense in the first experiment is still quite high, reaching 70% (see Table 1 p. 266). It is also noteworthy that the VPE sentences were judged to make sense less often even when the antecedent-trigger was parallel (89% vs 94% for VPA). Though they found no effect of parallelism on judgements to deep anaphors, it did affect judgement latencies to sentences that ‘made sense’ for both anaphor types. Specifically, higher latencies were observed when the antecedent was non-parallel. Although these results support the hypothesis that deep and surface anaphors are processed differently, Tanenhaus and Carlson acknowledge that the high ratio of mismatched VPE cases judged to make sense is unexpected, as is the finding (comparable to Murphy’s) that both anaphor types are affected by mismatch. This leads them

to suggest that the processing of mismatched cases of VPE may be analogous to the comprehension of ungrammatical sentences, involving some form of reconstruction process to arrive at the intended interpretation (this similar in spirit to later accounts such as [Arregui et al. 2006](#)). As for the finding that deep anaphors are also affected by mismatch, they argue that it may be linked to focus variations induced by passive or active antecedents which render the antecedent less accessible.

[Maurer et al. \(1995\)](#) re-examined the findings of [Tanenhaus and Carlson \(1990\)](#), and obtained rather more conclusive results than in the initial study. In a reanalysis of results from [Tanenhaus and Carlson's](#) first experiment on passive/active mismatches, separating full passives (with an overt agent) from short (agentless) ones, they found an effect of anaphor type with full passives but not short ones, and these results were replicated in their first experiment testing short passives only. However, in a second experiment testing full passives, they found an overall effect of parallelism on both deep and surface anaphors, similar to what [Tanenhaus and Carlson](#) had previously reported. More strikingly, they observed that parallelism effects were smaller than in the first experiment, a result they attribute to the awkwardness of passive sentences with *by someone* as the agent phrase, which they claim might have led subjects to adopt looser criteria for acceptability in their judgements.

[Roberts et al. \(2013\)](#) analysed the processing of mismatched VPA/VPE cases with both passive and nominal antecedents in an eye-tracking experiment. The choice of task was intended as an alternative to the judgement tasks employed in previous studies which, according to them, 'tap into later, interpretative processing, rather than into more immediate, incremental processing' (p. 33). They examined both the predictions made by [Sag and Hankamer](#) on the different processing of deep and surface anaphors, and the alternative claim defended by [Hardt \(1993\)](#) and others that VPE does not actually require syntactic identity, only semantic

identity. Their materials consisted of a lead-in and continuation sentence in addition to the antecedent and anaphoric clauses. Below are sample materials for the passive and nominal antecedent conditions:

- (199) It was snowing very heavily last night,
- a. but someone took the wood out to the shed. Tom told us that Sally did (it).
 - b. but the wood was taken out to the shed. Tom told us that Sally did (it).
- I was surprised that she was so kind. (1, p.46)
- (200) There is a great discrepancy of wealth in the US.
- a. Someone robbed CitiBank as an act of desperation. The police haven't figured out who did (it).
 - b. The robbery at CitiBank was an act of desperation. The police haven't figured out who did (it) .
- The bank remained open for the rest of the year. (12, p. 47)

Results were mixed, showing only partial evidence for a difference in the processing of deep and surface anaphors. Though there was an overall advantage for the VPA-parallel condition, [Roberts et al.](#) also observed that VPE took longer to process even in the parallel condition. As in previous analyses, they found that parallelism affected the processing of both constructions, rather than specifically VPE. They conclude that their results partly contradict [Sag and Hankamer's](#) proposals but do not support semantic accounts such as [Hardt](#) or [Dalrymple et al. \(1991\)](#). Instead, they are consistent with the recycling hypothesis proposed by [Arregui et al. \(2006\)](#).

5.1.3 Summary of findings and notes on materials

The studies reviewed so far provide mixed support for the central claim of a difference in processing between deep and surface anaphors, as only

some of them find evidence that the effect of mismatch is worse for VPE than it is for *do it* or other deep anaphors such as NCA. A more common finding was that lack of parallelism affected both anaphor types equally, contradicting the idea that deep anaphors are insensitive to antecedent form.

As pointed out by several authors, the choice of task and the materials used may in part explain the contrasting results obtained. Regarding the latter aspect, it is arguable that the antecedent sentences are often degraded independently of the matched/mismatched nature of the antecedent. Amongst other things, passive materials tend to have antecedent sentences which are themselves infelicitous to various degrees. [Tanenhaus and Carlson](#), as mentioned, pointed out the badness of long passives with *by someone* as the agent, yet the materials used by [Roberts et al.](#) include several sentences with indefinite pronouns in the *by* agent phrase, e.g., *by someone* in (201) below or *by almost everyone* in (202). Arguably, the active variant is more natural-sounding in (201), and even more so in (202). Even in those sentences with a full NP in the *by* phrase, the passive variant may still appear degraded, e.g., (203).

- (201) We couldn't have driven out due to the heavy snow. (But someone shovelled our driveway/But our driveway was shovelled by someone.) (A neighbour told us that Tom did/A neighbour told us that Tom did it.) We decided to hire Tom as a snow clearer. (ex. 4, p. 47)
- (202) Mrs. Brown threw a big party last night. (Almost everyone at the party ate far too much food/Far too much food was eaten by almost everyone at the party.) (Even Mary did/Even Mary did it), although she was a picky eater. Mrs. Brown was so happy that Mary liked her food. (ex. 16, p. 47)
- (203) Writing a novel that sells well is so hard. (But my brother Sam wrote a best-selling novel/But a best-selling novel was written by

my brother Sam.) (Our family was proud that he did/Our family was proud that he did it.) Now Sam is working on another one. (ex. 6, p. 47)

Of greater interest, however, is the acceptability of the VPE or *do it* materials independently of the antecedent clause. A number of the VPE *to* materials from [Tanenhaus and Carlson](#) are of dubious acceptability; for instance, one of the items for their Experiment 1, reproduced below in (204), is as follows:

- (204) a. An architect designed the elaborate conference room.
b. The elaborate conference room was designed by an architect.
c. He was paid a lot of money to do it
d. He was paid a lot of money to.

(ex. 5, p. 276)

(204c) seems intuitively more felicitous than (204d) regardless of whether the antecedent sentence is passive or active. In spite of this, (204d) can still be comprehended, as reflected in the sensicality judgements reported by [Tanenhaus and Carlson](#). Nevertheless, considering that the VPE versions are often so strongly degraded, it is unsurprising that they take longer to process in non-parallel conditions, if we take into account the general effect of mismatch on both anaphor types observed in their experimental results.

Problems such as these are at the heart of a number of studies placing greater emphasis on the role of discourse factors in the varying acceptability of mismatches. [Kehler \(2000, 2002\)](#) and [Kertz \(2010, 2013\)](#) showed that passive/active mismatches with VPE could be rendered more acceptable if general discourse conditions were respected.

In a study of nominal antecedents with VPE and *do it*, [Miller and Hemforth \(2014\)](#) elaborate on this line of work and further propose that there

are specific discourse conditions bearing on the use of VPE and *do it*, and that they will show reduced acceptability if these conditions are not respected. They explain that earlier accounts such as those discussed above did not take discourse conditions on VPE or *do it* into consideration, making their results harder to interpret. They also point to independent evidence that parallel structure facilitates processing not just of ellipsis or anaphoric expressions, but in general (Frazier et al., 1984, 2000; Frazier and Clifton, 2001; Dubey et al., 2005), which would explain the fact that both VPE and *do it* are sensitive to mismatch. Following Kertz (2008), they distinguish between two uses of VPE, ‘auxiliary-choice’ and ‘subject-choice’, each with their own set of conditions. Aux-choice VPE is characterised by coreferent subjects in the antecedent and VPE sentence as well as prosodic stress on the auxiliary. This usage requires that the discourse context contain a highly salient choice between members of a jointly exhaustive set of alternative situations, and further that the VPE sentence be ‘strictly limited to selecting one member of that set’. (p. 4). ‘Subject-choice’ VPE involves different subjects in the antecedent and VPE sentences, and stress on the subject if it is a pronoun. It also requires that a particular property be salient in context, with the VPE sentence limiting itself to ‘identifying something or someone having that property.’ (p. 4).

(205) A.— Does he shop in women’s? B.— He DOESN’T/He DOES.

[Aux-choice]

(206) A.—Mark shops in women’s. B.—ANDY does too. [Subj-choice]

Miller and Hemforth (2014) further show that sentences violating the proposed conditions are less natural, even if the antecedent is parallel. Importantly, they also point out that *do it* is intuitively much less acceptable than VPE in Aux-choice contexts, a preference confirmed by evidence from an acceptability experiment showing reduced acceptability *do it* compared to VPE in answers to polar questions. Their nominal antecedent materials are based on the attested example below, which involves AUX-choice

VPE. They explain that the noun *survival* is what they call a ‘polar noun’, by which is meant that it is interpreted here as a ‘concealed question’ with the meaning ‘whether or not Mubarak will survive’, making the alternative {Mubarak will/will not survive} salient. Since moreover the VPE sentence simply chooses a branch of the alternative, without more, the proposed conditions are met.

- (207) Mubarak’s survival is impossible to predict and, even if he does/
#does it, his plan to make his son his heir apparent is now in
serious jeopardy. [COCA: CBS Evening News] (ex 1, p. 3)

[Miller and Hemforth](#) investigated the acceptability of nominal antecedents such as these in a judgement task comparing verbal and nominal antecedents of VPE and *do it* in contexts with or without a salient polar alternative (labelled +/–Alt). Results confirmed the preference for VPE in +Alt contexts and *do it* in –Alt contexts. However, VP antecedents were judged more acceptable than nominal ones, especially with VPE. Importantly, the *Do/N/+Alt* condition was still rated quite high (5.15 on a 7-point scale), and higher than the *Do it/V/+Alt* or *Do it/N/+Alt* conditions, meaning that sentences with mismatched VPE that respected the proposed discourse constraints were judged more acceptable than those of *do it* were the conditions were not respected.

While these results confirm that mismatched VPE can be rendered more acceptable in certain contexts, they still show a penalty for nominal antecedents as compared to verbal ones. [Miller and Hemforth](#) attribute this effect to the higher processing cost of the nominal antecedents due to their particular semantics, specifically because they require accommodation in order to be interpreted as questions. In a separate experiment testing the antecedent sentences only, they found that nominal antecedents were judged less acceptable than verbal ones in the +Alt context, but higher in the –Alt context, suggesting polar nouns are indeed harder to process. Additionally, they claim, following [Kertz \(2013\)](#), that lack of parallelism

further slows down processing by making it harder to ascertain semantic identity with the antecedent.

5.2 Antecedent mismatch with VPAs

Experimental studies offer valuable insights into the processing of mismatched cases of VPE/VPA, but say little about the kinds of mismatches that are actually attested. The specific case of *do this/that* is never discussed, presumably because it is assumed that, like *do it*, they allow mismatched antecedent-triggers in all cases. This section will re-examine the question of mismatch with VPA based on extensive corpus data, discussing passive as well as nominal antecedent-triggers

Methods

Since the BNC is not annotated with anaphoric relations, there is no way to know what the antecedent-trigger is, or even if there is one in the first place, without looking at the extended context of the VPA. It is also impossible to search directly for specific structures occurring as antecedent-triggers, making it difficult to collect occurrences of mismatched or exophoric VPAs. In Chapter 4, these limitations were circumvented by making use of heuristic strategies to retrieve non-agentive antecedents. In the case at hand, this is not possible, since there are no contextual features that make it more likely to have, say, a passive VP or an NP antecedent-trigger. The analysis will therefore be limited to those occurrences found in already-available data from the BNC: the initial sample of 900 occurrences discussed in Chapter 2, along with the sample used for the study of agentivity in Chapter 4, which contains 747 occurrences.

5.2.1 Mismatch and inference

As suggested at the beginning of this chapter, a central distinction can be made between cases of syntactic mismatch, where the structure of the antecedent-trigger does not match that of the VPA, and those where the antecedent is inferred either from previous context, or from the speech situation (exophora). Also included in this category are examples like (208), where *do it* is not understood as ‘catch fire’ but as ‘set fire to the village’. Here, although there clearly is an antecedent-trigger, it does not supply the full interpretation of the VPA, requiring further inference to retrieve the actual antecedent.

(208) Why did our village catch fire? — The gods did it. (CET)

Table 5.1 shows the frequency of the different possible structures for the antecedent-trigger—active or otherwise—as well as inferred antecedents. ‘NA’ occurrences are those where the VPA occurs in a passage of quoted direct speech that does not include the antecedent sentence, so that the structure of the antecedent-trigger (if there is one) cannot be determined with certainty.

The vast majority of occurrences in the corpus have an active, hence syntactically parallel antecedent-trigger. Inferred antecedents are the second most frequent case. Among mismatched cases, passive VPs are more frequent than NPs. The interaction between VPA choice and structure of the antecedent-trigger was found to be significant ($p < 0.05$): this may be due to the lower frequency of NPs with *do this* and passives with *do that*.

Table 5.2 shows the distribution of matched and mismatched cases across VPAs for those occurrences where there is an antecedent-trigger (this does not include cases of inference of the type shown in 208). As was already apparent from the data in Table 5.1, having a parallel antecedent-trigger is the most typical situation. The frequency of mismatches does not vary significantly across VPAs ($p = 0.4$), suggesting that all of them

5.2. ANTECEDENT MISMATCH WITH VPAS

allow mismatch to a similar degree.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Active	566 (78.6%)	517 (83.1%)	250 (82%)	1333 (80.9%)
Inference	38 (5.3%)	43 (6.9%)	23 (7.5%)	104 (6.3%)
Passive	32 (4.4%)	37 (5.9%)	8 (2.6%)	77 (4.7%)
NA	34 (4.7%)	13 (2.1%)	9 (3%)	56 (3.4%)
Pronoun	27 (3.8%)	7 (1.1%)	6 (2%)	40 (2.4%)
NP	23 (3.2%)	5 (0.8%)	9 (3%)	37 (2.2%)
TOTAL	720	622	720	1647

Table 5.1: Active, mismatched and inferred antecedents

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Match	586 (91.4%)	517 (92.5%)	256 (93.8%)	1359 (92.3%)
Mismatch	54 (8.6%)	42 (7.5%)	17 (6.2%)	114 (7.7%)
TOTAL	641	559	720	1473

Table 5.2: Matched (active) and mismatched antecedent-triggers

5.2.1.1 Mismatch and agentivity

Chapter 4 showed that non-agentive VPAs are infrequent but just as acceptable as agentive uses, and that in most attested examples the subject is a non-volitional cause of the action. It appears that non-agentive VPAs also allow mismatch, since as can be seen from Table 5.3, the corpus contains six occurrences that have both a mismatched antecedent-trigger and a non-agentive antecedent. In three of them the trigger is a passive VP (209), and in the remaining three it is an NP that denotes a non-agentive process, such as *the generation of helium-3...* in (210). Only *do it* and *do this* are attested in non-agentive mismatches, although it was reported in Chapter 4 that *do that* occurs significantly more often in non-agentive uses than *do it/this*. This observation is presumably a consequence of the very

small number of mismatches with *do that* (just 17 occurrences in total) together with the fact that non-agentive uses are also infrequent.

- (209) In a remarkable demonstration of this, Johansson and Maas took cine-films of human figures with lights attached to their joints—knees, elbows, and so on—but otherwise in the dark. If the figures are moving, they are perceived as what they are. To do this, we must have in our heads a precise but flexible model of how people move. (AE7)
- (210) [Citation in science book] [They] attribute this [the presence of Helium in volcanic gases] to primordial helium...being released...I attribute it to the generation of helium-3 by the fusion of deuterium and tritium by something or other. I would guess that the ‘something or other’ are negatively charged ions of fluorine, chlorine and maybe oxygen. Maybe under pressure free electrons do it. (CER)

One question arising from such examples is whether mismatch is affected by the agentivity of the antecedent. The data in Table 5.3 suggest this is the case: a lower frequency of mismatches is found in non-agentive than agentive cases. This effect was found to be significant ($p < 0.05$). It may be that the combination of mismatch and non-agentive antecedent makes the VPA more difficult to process, though this is not apparent from the two examples given.

	Match	Mismatch	TOTAL
Agentive	1198 (88.2%)	108 (94.7%)	1306 (88.7%)
Non-agentive	161 (11.8%)	6 (5.3%)	167 (11.3%)
TOTAL	114	1473	1473

Table 5.3: Frequency of matched and mismatched antecedent-triggers according to the agentivity of the antecedent

5.2.2 Passives

Since the queries used when building the corpus searched only for active occurrences of VPA, all examples discussed here involve mismatches between a passive trigger and an active VPA. Since VPAs can be passivised, the reverse pattern is also attested in the BNC, as already pointed out in Chapter 3. The corpus does however contain one occurrence of passive *do this* with an active trigger (211), which was incidentally found in the broader discourse context of an occurrence of active *do it*.

- (211) Ms Burke said that under her care, Profile provided about 3,000 telephone numbers a day to customers undertaking planning applications. This was done using British Telecommunications Plc's Phonebase system to access the directory enquiries database. But, because Phonebase is notoriously difficult to access and the charging procedure is complex, Ms Burke and her husband decided there must be a better way of doing it. (CPW)

5.2.2.1 Presence or absence of a passive agent

Following Biber et al. (1999), we refer to passives followed by an agent *by*-phrase as 'long passives', and to agentless ones as 'short passives'. Most passive-active mismatches in the corpus involve short passives (see Table 5.4). This reflects the general tendency for English described by Biber et al., who write that 'short passives are predominant in all syntactic positions' (p. 937; see also their table 11.9, p. 938, for further details).

TYPE	
Short passive	72 (93.5%)
Long passive	5 (6.5%)
TOTAL	77

Table 5.4: Short and long passives as VPA antecedent-triggers

This is in contrast to the rather frequent use of long passives in experimental studies of mismatch; for instance, [Maurer et al. \(1995\)](#) point out that the materials in [Tanenhaus and Carlson \(1990\)](#) include 9 long passives and 10 short ones, adding that many of the long passive sentences were ‘somewhat awkward’ (p. 4). This problem has already been observed with the materials from [Roberts et al. \(2013\)](#).

The six occurrences that have a long passive in the antecedent-trigger include four of *do it* and two of *do that*. In one of the *do it* cases, the VPA is embedded inside the passive agent (212). Note that in this case the *by*-phrase brings essential new information, since it has already been stated that the tables are carved; deleting the agent would make the sentence infelicitously repetitive. In other examples, the agent of the passive sentence is often mentioned again in the VPA clause (213)-(214).

- (212) I remember, again (laugh) a few years ago, visiting India and buying one or two beautifully carved tables, which had obviously been carved by an individual spending quite a lot of time doing it, and I was impressed at that stage, rather naively perhaps, that if in fact I’d bought a plain table, an uncarved one, it would have cost me about ten times as much [...] (KRG)
- (213) ‘Constable Bewman here pointed out that each guest had their plate handed to them by Edith but I can’t see how that would give the murderer any scope.’ Henry frowned. ‘She would have done it in a preordained way, of course,’ mused Henry. (CDN)
- (214) It [a letter] addressed air quality they criticize the threat to air quality in their letter and that hasn’t been addressed by the NRA it’s beyond the NRA’s permit to to do that. (HMP)

In examples with short passives, the agent may be omitted because it is unknown (215)–(216), obvious in context (217), unspecific or generic (218). Most often, the subject of the VPA is identified with the passive

5.2. ANTECEDENT MISMATCH WITH VPAS

agent, whether or not it is known. In (217) the speaker and addressee, referred to by *we*, are the ones responsible for making the shoes. In (215)–(216) the unknown agent is picked up by interrogative *who*, and likewise *you* in (218) is interpreted as generic in the same way as the unexpressed agent.

- (215) It was clear that funds were being embezzled but who could be doing it? Could it possibly be the accountant Edward Morris?
- (216) But, if he was murdered, the important thing is not who did it but why. If it was a personal vendetta. (B3J)
- (217) ‘Don’t talk, Will, there’s too much work, six pairs of shoes to be made by the weekend! I don’t know how we’re going to do it, even if we work flat out every day.’
- (218) I mean, is that an attitude that that that should be changed? And if so, how do you do it? (FLD)

It follows from this that the subject of the VPA sentence is usually coreferent with the implicit or explicit agent of the passive sentence. This is the case in 77.9% of occurrences (see Table 5.5).

COREFERENCE OF SUBJECTS	
Same	60 (77.9%)
Different	17 (22.1%)
TOTAL	78

Table 5.5: Coreference of subjects between the antecedent and VPA clause with passive antecedent-triggers

Cases of non-coreference between the passive agent and the subject of the VPA occur primarily with short passives; there are only two examples of mismatch between an overt *by*-agent and the subject of the VPA sentence. In most such cases the agent is interpreted as generic, and therefore includes the VPA subject; for instance, in (219), the ban on record-keeping

applies to all organisations including CND. Alternatively, if the agent unknown, the VPA sentence may serve to identify it (220). (221) is a rare example of an actual contrast between the implied agent of the passive sentence and the subject of the VPA clause, with a comparison between how distances might be estimated by the human brain as opposed to a computer.

(219) ‘It’s a traditional cry of the left, that no records should be kept on individuals, but CND have found an excuse to do it,’ Nicholas Perry, co-author of the booklet told New Scientist.’ (B7J)

(220) Why have we been flooded? – The gods did it. (CET)

(221) We do not know how cognitive maps are stored in the brain. However, given that such maps exist, I find it hard to believe that distances could be estimated by a method analogous to using Pythagoras’ theorem. A computer programmer would almost certainly get a computer to do it that way, but I doubt whether the brain does. (AE7)

(222) shows the only case where the subjects can be construed as disjoint in reference. It exhibits the coordination of a long and short passive VP, with the agent of the first VP being also understood as the agent of the second. Although *we* and *the government* are clearly not coreferent, the shift to a first-person plural subject is motivated by the generic interpretation of the *do that* sentence (as shown by the remnant complement *to someone*), and the generalisation from the authorities’ attitude towards Koresh to society’s treatment of criminals in general. It thus seems that this example should be interpreted as a case where the subject of the VPA includes the subject of the antecedent, rather than a case of disjoint reference.

(222) — Koresh has been demonized by the government, and his followers ridiculed as cultists. They became the object of public outrage

and loathing. And very few stopped to think about what had happened.

— The true test of a free society is not in how it treats its best citizens, but in how it treats its worst, its most despised. And if we can do that to someone because, well they're religious nuts, (unclear) they're they're a little bit different. Or they're a lot different. Then we can do it to you or me or anyone else, and that troubles me. (HE3)

The rarity of non-coreferent subjects in occurrences of mismatch is in keeping with the constraint proposed by [Kertz \(2013\)](#) that contrastive topics prefer to be realised as sentence subjects in ellipsis contexts as well as in general. As with long and short passives, the experimental studies discussed earlier often overlook the role of identity between subjects, mixing sentences where they are coreferent (such as (204) above where *he* refers to the architect) and others where they are not (223):

- (223) a. Somebody had to paint the garage.
 b. The garage had to be painted.
 c. Finally my younger sister Carol agreed to do it.
 d. Finally my younger sister Carol agreed to.
 ([Tanenhaus and Carlson, 1990](#), ex. 18, p. 277)

5.2.2.2 Alternation with VPE

65 of the 77 passive/active mismatches have a non-finite VPA, and in 5 of the finite occurrences, the VPA is followed by an adjunct. This means that VPE with auxiliary *do* is most often ungrammatical or at least dispreferred. In finite adjunct-less cases, the acceptability of VPE depends on whether the context contains a salient choice between polar alternatives, as proposed by [Miller and Hemforth \(2014\)](#). (224) is a case of auxiliary-choice (the *if*-clause creates an alternative between orthogonalising the columns

or not) and VPE may be substituted for *do this*. In (225), by contrast, the passive VP occurs in a relative clause and the antecedent is not asserted, so that the alternative is not salient.

(224) However, it is probably sufficient if the columns are orthogonalised at, say, every fourth iteration. If we do this/do, the procedure is as follows. (EWW)

(225) Yet these million or so termites build their equivalent working in a coordinated way in total darkness, each blind, tiny-brained insect knowing exactly where it has to place its pellets of mud to produce nurseries, supporting pillars, living chambers, gardens, flues, defensive walls — and that extraordinary spiral cooling vane. As with so many of the buildings constructed by animal architects, we really have very little idea how they do it/#do. (F9F)

All finite cases other than those given conform to the subject-choice usage of VPE described by Miller and Hemforth, insofar as the passive agent is unexpressed and the point of the VPA sentence is to identify that agent (226), or at least ask about his/her identity as in (227). Contrary to the auxiliary-choice cases, VPE is less felicitous here:

(226) Why have we been flooded? – The gods did it. (CET)

(227) I hear your husband was blackballed when he tried to join the Country Club. Wonder who did that#did? (A0D)

Where the VPA is non-finite, VPE may still occur if it follows an auxiliary or *to*, and the relevant discourse conditions obtain. Thus it is felicitous in (228), which satisfies the requirements for aux-choice, but not (229), where the requirements are not met.

(228) Sun Microsystems Inc has denied that it has plans to develop a Sparc implementation of Microsoft Corp's NT operating system, as reported last week in the US paper Open Systems Today: the

5.2. ANTECEDENT MISMATCH WITH VPAS

paper quoted Sun president Scott McNealy as saying ‘NT needs to be ported to Sparc—if they don’t (do it), they are missing the boat’ (CNJ)

- (229) Fig. 5 also shows how the board is fitted to the panel—but don’t #(do this) just yet. (C91)

Note that given an appropriate context, VPE may become felicitous even in non-finite cases where no auxiliary is present, provided that the anaphoric sentence is altered in such a way as to require a finite VP. For instance, **by doing* alone is impossible in (230), but the beginning of the anaphoric sentence may be felicitously reworded to *If we do*.

- (230) In the creation of effective multimedia the professions of information scientist, designer, computer programmer, systems analyst, film maker, financier, and publisher must somehow be drawn together, either in a single extraordinary person or into a well-managed and effective creative team. By doing this we may begin to see products that people will want to buy and use. (HRD)

5.2.3 Nominal antecedent-triggers

NP antecedents have been studied less often than passive/active mismatches, and mainly in relation to ellipsis. Hardt (1993) provides four attested examples and discusses other constructed ones. Several psycholinguistic studies have investigated the acceptability of different types of nominal antecedent-triggers with VPE and/or *do it*, among which nominal gerunds (Arregui et al., 2006) and various subtypes of deverbal nouns (Roberts et al., 2013; Miller and Hemforth, 2014). As with passive/active mismatches, there are no corpus-based studies of NP antecedent-triggers with VPAs.

5.2.3.1 Conditions on nominal antecedent-triggers

Two general conditions may be proposed for nominal antecedent-triggers. First, they should either denote an action, or at least be sufficiently transparent to allow an action meaning to be inferred (see [Ward and Kehler 2005](#)). Example (240) is a case with an action-denoting noun (*passage*), where the VPA is straightforwardly interpreted as ‘pass from my armchair to the kitchen’. (232) involves an agent nominal, *songwriter*, and *do that* is understood as *write songs*.

(231) I want now to turn to a less philosophical question about my passage from my armchair to the kitchen. How do I actually do it? Figure 10 shows an (imaginary) map of my house. A is the armchair and E the gas stove. The dotted line A-C-D—E represents the only sensible route. The task of going direct from A to C, if I can see C, presents no special difficulty; this is the kind of problem solved by the students of animal orientation. (AE7)

(232) So what’s what’s your occupation Martin, are you a songwriter?
Erm well not as such as a songwriter, I mean I do that yes but I’m my occupation I’m a self-employed window cleaner. (HMD)

If an NP is too opaque to infer an action meaning, the VP will be infelicitous, as already observed by [Ward and Kehler](#) for *do so* with examples like (233), where *computer* cannot lead to interpreting *do so* as ‘compute’.

(233) # My computer does so faster than yours [= compute] (ex. 39)

The second condition is that NP triggers with *do it/this/that* should preferably denote agentive actions. Thus, replacing *computer* with *calculator* in [Ward and Kehler](#)’s example (see 234) is still infelicitous despite being more transparent, since the antecedent is not agentive. As shown in Chapter 4, VPAs may also have non-agentive uses, and the corpus does in fact contain occurrences where the nominal antecedent-trigger denotes a non-agentive process.

(234) # My calculator does it/this/that faster than yours. [= calculate]

5.2.3.2 Overview of corpus data

NP antecedent-triggers may be classified according to the morphology of the head noun, specifically whether it is a deverbal noun, such as an agent nominal in (235), a nominal gerund (236), or a non-derived noun, e.g., *member* (237).

(235) ‘What are your plans, then? I suppose if you’re a writer you can do it anywhere. Must be wonderful.’ (CKB)

(236) But the most pure expression of capital—capitalism is the the harnessing of private money to start private businesses, and a main concern of the institute of course, is that there is such a shortage potentially of capital, in the system, to do that. (HMH)

(237) You know, if I’m president of the board of trade, I’m a loyal member of John Major’s cabinet. I intend to go on doing that as long as he wants me. (K6A)

Additionally, the antecedent-trigger may be a compound noun (238) or a coordinate NP (239).

(238) Tool using was thought to be the ‘key’ to understanding ourselves. It is now apparent that too much weight has been attached to tool using. To the animals that do it, there is nothing particularly special about using tools: it is a piece of behaviour much like any other that the animal performs. (CJ3)

(239) At the end of the Liberal Democrats Concert, er Concert (laugh) Conference we had all sorts of singing and er jollifications, do the Greens do this in, on their final day? (KRT)

For the purposes of corpus annotation, we only retain the morphological distinction proposed above, without treating compounds and coordi-

nation as categories of their own. Deverbal nouns also include examples of conversion such as *release* in (240), as well as non-derived ones.

- (240) A further sign of pressure from Islamic radicals came today from Hussein Mussawi, in the middle here, a senior Shiite leader closer to Hezbollah fundamentalists, he opposed the release of another American and questioned the loyalty of those doing it. (KRU)

Table 5.6 shows the distribution of the different nominal antecedent-triggers (due to the small size of the sample, we will not consider differences in frequency between VPAs). Most of the time, the head noun is either a deverbal noun or a nominal gerund.

HEAD NOUN	
Deverbal noun	23 (62.2%)
Nominal gerund	7 (18.9%)
Other noun	7 (18.9%)
TOTAL	39

Table 5.6: Types of NP antecedent-triggers

5.2.3.3 Deverbal nouns

Deverbal nouns usually denote an action or the result of that action or, in the case of agentive nominals, the agent of the action. Amongst action-denoting nouns, conversions show the highest degree of transparency, as they are morphologically identical to the verb, making it relatively easy to retrieve the antecedent. For instance, in (241) below, *do this* is interpreted straightforwardly as ‘support OMG’s Common Object Request Broker Architecture’.

- (241) At Object World, NeXT Computer Inc is expected to announce its membership of the Object Management Group. IBM, already a

5.2. ANTECEDENT MISMATCH WITH VPAS

member, is to announce its support of OMG's Common Object Request Broker Architecture, CORBA. It intends to do this by modifying the object interface definition language in its System Object Model (SOM). for OS/2 2.0 to support CORBA's interface definition language. (CTN)

Example (242) provides a case where the noun has a result interpretation: *measurement* may of course refer to the process of measuring, but here, it occurs as the complement of *need* and denotes the data obtained as a result of measuring. It is therefore clearly stative. The VPA does not preserve this meaning, however, and is interpreted dynamically as 'measure the energies of the electron'.

(242) They were therefore still needing the measurements of the energies of the neutrons, which were necessary as a proof that they were indeed neutrons produced by dd fusion and not somehow spurious. Their first piece of fortune in mid-March was apparently finding a way to do this. (CER)

Agent nominals are a very frequent category within the deverbal set (though none are attested with *do this*). In the most simple case, the antecedent can be inferred from the verb from which the noun is derived, e.g., *writer*>*write* (243). However, the interpretation may be less straightforward and require further inference, for instance *do that* in (244) does not mean simply 'entertain' but rather something along the lines of 'act as a (professional) entertainer'.

(243) 'What are your plans, then? I suppose if you're a writer you can do it anywhere. Must be wonderful.' (CKB)

(244) I was happy as pop singer, the last couple of years I wasn't happy when I ceased to be a pop singer and I was er headed on a road of being an entertainer. I could not be happy doing that, but as a pop singer I was in my element, because really all I d- I didn't

even think of it as singing, I thought of it, when I performed on stage as a pop singer I just thought of of it really of making love to the audience. (KGH)

Agent nouns typically occur as predicates in the antecedent-trigger; (245) shows a less common pattern where the antecedent-trigger occurs in the subject NP.

(245) The joyrider that we interviewed, he was fairly upset and he wished that he'd never done it in the first place. (FXT)

Only one of the agent nouns is non-agentive (246), but the 'success' meaning along with the manner adjunct both favour *do it*.

(246) I've always looked upon the horse as a nice horse not as the winner of an Arkle. And he did it the right way, he did it the hard way [...] (KRM)

Like [Ward and Kehler](#) in their corpus study on *do so*, we found no occurrences with nouns denoting inanimate entities. The acceptability of such nouns therefore remains an open question. The corpus also does not contain any examples of nouns occurring in the '+Alt' context described by [Miller and Hemforth \(2014\)](#), where the noun expresses a concealed question that makes a choice of alternative salient (for instance, in (247), whether or not Mubarak will survive). This is consistent with their experimental results, which show that +Alt contexts favour VPE over *do it*, as well as the claim by [Miller and Pullum \(2014\)](#) that *do this/that* are also dispreferred in such cases.

(247) Mubarak's survival is impossible to predict and, even if he does/ #does it, his plan to make his son his heir apparent is now in serious jeopardy. [COCA: CBS Evening News] (ex 1, p. 3)

5.2.3.4 Nominal gerunds

Nominal gerunds are similar to conversions with respect to their degree of transparency: they are morphologically identical to the *-ing* form of the verb and denote an action, making the antecedent relatively accessible, as with *dieting* > *diet* in (248). They are in fact close to verbal gerunds like (249).

- (248) Gurin and Bennett provide a scientific analysis of dieting and why most people shouldn't do it. (B7M)
- (249) Escaping is bad enough but doing it in the middle of the night is inexcusable. (AMB)

5.2.3.5 Other nominal antecedent-triggers

As pointed out above, this category includes both derived and non-derived nouns. A further distinction can be made between nouns that denote actions, e.g., *sex* in (250) (the fact that it refers to an action here is confirmed by the fact that it is the subject of *happened*) and those that do not, e.g., *surrogate mother* (250).

- (250) Donal Hickey at the University of Ottawa suggested a decade ago that sex – by which he meant the entirely puzzling need to mix the genes of different individuals while reproducing – first happened because certain genes selfishly forced their possessors to do it. A gene is stuck in the same lineage without sex. (AKF)
- (251) I'd like to say I was a surrogate mother and I wasn't paid a penny for doing it! But I did it through the love, fo–, that I had for the couple, and that because they had waited sixteen year before they eventually found out they couldn't have children! (FLG)

With action-denoting nouns, as already observed, the antecedent is usually quite accessible. Non-actional nouns, on the other hand, require

some further inference. In all cases the antecedent-trigger occurs as a predicate and is therefore stative, but it is reinterpreted as an action in the VPA clause, for instance *do it* in (251) means roughly ‘act as a surrogate mother’. This is similar to what was observed with some agent nouns above.

5.3 Exophora and inferred antecedents

This section will be concerned with what in traditional accounts has been described as ‘antecedentless anaphora’, i.e., where the interpretation of the VPA does not rely on a well-defined antecedent-trigger, but is instead achieved via inference from the linguistic or extralinguistic context.

(252) is an example of the type commonly discussed in the literature on anaphora (albeit not with VPAs) where *do that* refers exophorically to the action the addressee was performing just before being interrupted by the speaker, i.e., brushing her hair. Since this action was still ongoing shortly before the exchange, it is sufficiently salient to be understood as the antecedent of *do that* without any explicit description of the action. (253) similarly lacks an antecedent-trigger in the sense that no single structure can be identified as the source of the interpretation. Rather, the antecedent is retrieved ‘incrementally’ based on ‘retiring to various parts of the country’ and ‘to the pleasant places’. Since the extralinguistic context does not in any way contribute to this interpretation, the VPA cannot be considered as exophoric.

(252) I caught her trying to brush her hair the other day. She knows she can’t do that! She can’t lift her arms up. I said, ‘Give that here,’ and got the brush off her. ‘I’m supposed to do that,’ I said. But then she went all moody. (ACB)

(253) They’re retiring to various parts of the country, down here onto the Sussex coast and on the whole coast into the South West—to

5.3. EXOPHORA AND INFERRED ANTECEDENTS

the pleasant places if they can afford to do it [...] (KRE)

In addition to ‘antecedentless’ cases like the above, we will consider occurrences of VPA where an antecedent-trigger can be identified but is either very distant or does not provide a full interpretation of the anaphor, for instance (254) where *do it* means not ‘die’ but ‘kill O’Shaughnessy’.

(254) It is terrible that Alison O’Shaughnessy died in such a brutal way and her family has our heartfelt sympathy. But it wasn’t our girls that did it and I am going to prove it.’ (CEK)

Given [Hankamer and Sag \(1976\)](#)’s observation that *do it* (as well as, presumably, *do this/that*) may be used without a linguistic antecedent, examples like the above are expected to be acceptable so long as the extralinguistic context provides sufficient information to recover the antecedent. As such, they do not pose a challenge to the deep vs. surface anaphora distinction in the way that mismatched or exophoric examples of VPE do. Consequently, there has been little investigation into the exact conditions under which VPAs may be used in this way, beyond the requirement of an accessible antecedent. As discussed in Chapter 3, [Huddleston and Pullum \(2002, p. 1532\)](#) offer some comments on exophoric *do this/that*, explaining how they might refer to an action being performed by the speaker (*do this*) or the addressee, but make no mention of *do it*, which appears to be impossible in the examples given (e.g., saying #*Don’t do it!* rather than *Don’t do that!* when asking the addressee to stop what he/she is doing). [Miller \(2011\)](#) suggests that exophoric *do it* requires a more salient antecedent than *do this/that*, and further that *do that* has the lowest requirement of all, as illustrated in the following invented example where *do that* is preferred over *do it* as well as *do this*:

(255) [addressee is making an annoying noise tapping his pencil against the table while speaker is trying to concentrate; no mention of this noise has occurred in the discourse context]

Stop doing that! /#Stop doing it/this. (ex. 12, p. 90)

Cases of inference like (252)–(254) have also been rarely discussed, though Miller reports that several such examples of *do it* are found in the COCA.

In the summary of corpus data in Section (5.2), exophoric uses and other ‘antecedentless’ cases were collectively labelled as ‘inference’, since the antecedent needs to be inferred from the extralinguistic situation without reference to the discourse, or through the processing of previous discourse, possibly by means of an antecedent-trigger that provides a partial interpretation of the antecedent. In what follows, we will consider occurrences of the first type as exophora, while using the term ‘discourse inference’ to describe those where the linguistic context contributes more or less directly to accessing an antecedent. The frequency of these two categories in the corpus is shown in Table 5.7.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Disc. inference	33 (86.8%)	27 (62.8%)	13 (56.5%)	73 (70.2%)
Exophoric	5 (13.2%)	16 (37.2%)	10 (43.5%)	31 (29.8%)
TOTAL	38	43	23	104

Table 5.7: VPAs with inferred antecedents: exophora and discourse inference

The association between VPA choice and the discourse-inferred or exophoric nature of the antecedent was found to be significant ($p < 0.05$). Moreover, a significant difference was observed between *do it* and *do this/that* ($p < 0.05$ for both), but not *do this/that* ($p = 0.6$). This suggests that *do this* and *do that* are more frequent in exophoric uses.

5.3.1 Exophoric VPAs

As [Cornish \(1996\)](#) points out, exophora has traditionally been conceived of as reference to entities present in the extralinguistic context—meaning, in the case of VPAs, that the action is happening at the time of speaking. [Cornish](#) argues, however, that exophora does not in fact require the referent to be present in the speech situation, but rather that the antecedent be salient in context. The corpus contains several examples where the exophoric VPA does not refer to an ongoing action. For instance, *do this* in (256) refers to the action of breaking the door, which occurred before the conversation, but is inferred from the current situation of the door being damaged.

- (256) Sam disgustedly fingered the splintered door frame. ‘Did you sod-ding do this?’ he demanded.
‘It wasn’t locked.’
‘It was,’ I said. ‘With no key in sight.’
‘The key was in the keyhole on the inside.’
(ADY)

All attested occurrences of exophoric VPAs are found in spoken discourse, most of them in fictional dialogues. The spoken sample contains two occurrences in radio broadcasts, one in a university lecture, and a third in spontaneous conversation.

In most of the fiction examples, the narration provides either an explicit description the action, such as Masha’s attempt to resuscitate Rozanov in (257), or at least provides contextual clues from which the reader may infer the antecedent, as in (258) where *do this* refers to the unscrewing of the board. The point of such descriptions is to provide the reader with information about the speech situation to which he/she does not otherwise have access; in other words, the VPA would be uninterpretable if relying on the dialogue alone. Used in this way, *do it/this/that* are exophoric only

on the level of the dialogue, but not from the reader's perspective.

- (257) Masha, kneeling, was pressing her mouth to Rozanov's, attempting to force life into his body. At last I got through, demanded an ambulance, and had to ask Masha for the number of the room. She pulled back from him, sitting on her haunches. 'It's no good. I'm not even sure if I'm doing it right, or if it's the right thing to do.' (AEO)
- (258) When one of his [a teacher's] blackboards fell onto his foot because Endill had taken the screws from it, he refused to let the class leave until he had discovered the culprit. 'Who did this?' he shouted. 'Tell me now or you'll all have three lengths of the corridor.' Everyone looked at each other, but no-one spoke. (AMB)

More rarely, it is the dialogue rather than the narration which allows the antecedent to be inferred from the fictional speech situation. In (259), *do this* refers to Bernard's offer of accommodation in other words, to the speech act realised by his first and third utterances. The fact that the offer is reiterated just before the *do this* sentence together with the use of the progressive (*Why are you doing this?*) help both the reader and addressee to access the intended antecedent.

- (259) 'Look, I've got a spare room. You can use it if you want.'
'Thanks, but...' she trailed off with an awkward shrug. 'I mean, I don't even know you.'
'Likewise,' Bernard replied. He bit his lip thoughtfully.
'I'll tell you what. Call your friends and see if they can put you up for the night. If they can't you can either stay at the flat or else I'll give you some money and drop you off at a hotel.'
'Why are you doing this?'
'My father raised me. I never knew my mother. He was the only

5.3. EXOPHORA AND INFERRED ANTECEDENTS

family I had.’
(EF1)

A similar use of *do this* referring to the current speech act is found in the spoken BNC (260). In this case it refers to the speaker’s prediction as to whether or not the IRA will bomb Dublin. This interpretation is facilitated by the two previous sentences, in which the speaker expresses his refusal to make a ‘prophecy’ in reply to the interviewer’s question. Note that *do this* also occurs with the progressive here.

(260) Is [the IRA] bombing Dublin a real possibility?

Erm of course there’s a real possibility. I mean the one thing that has happened in recent times, and I don’t like prophecies so, I don’t wish to be a part of one, let’s be clear about that, I’m doing this as an analyst. Erm the loyalist paramilitaries have become much more sophisticated than they’ve ever been in their history. (HV6)

As noted above, exophoric VPAs are relatively infrequent in the spoken BNC (no examples of *do it* are found in our sample), meaning that the speech situation is often insufficient for the antecedent to be recovered without an actual linguistic description of it in previous dialogue. Examples taken from film scripts in the ‘Movies’ section of the Corpus of Contemporary American English (COCA) offer an interesting point of comparison (there is no equivalent section in the BNC). Like fiction works, scripts contain a mix of narration and dialogue, but the narration is unavailable and in any case unnecessary to the audience, who has access to a visual representation of the events in the film. Examples of *do this/that* are commonly attested (261)–(262), but occurrences of *do it* (263) are harder to come by.

(261) Pauline marches toward Malcolm. Malcolm wears the safety goggles and revs the chainsaw.

PAULINE Are you able to do this yourself?

CHAPTER 5. ANTECEDENT MISMATCH

MALCOLM Pauline, do you want me to cut it down or not?
(COCA, FIC: Mov: *Margot's Wedding*)

- (262) She (Cecilia) puts her cigarette between her lips and bends to pick up the vase, preparing to dunk it in the fountain, having first taken out the flowers and laid them on the step.

ROBBIE Let me do that.

CECILIA I'm all right, thanks.

But ROBBIE persists, reaching for the vase.

(COCA, FIC: Mov: *Atonement*)

- (263) The Sheriff throws his jacket on, starts walking.

SHERIFF RYAN You've lost your mind.

MIRANDA Don't fucking move.

The click! of a safety being released makes him pause. He turns to look at her. A tense beat.

PETE (O[ff].S[reen].) Miranda, don't do it !

She glances at Pete's pleading face and the room around her: cops everywhere. All guns trained on her.

(COCA: Fic: Mov: *Gothica*)

To a certain extent, the choice between *do this* and *do that* in exophoric contexts is determined by whether or not the speaker is performing the action. This distinction was already observed with [Huddleston and Pullum](#)'s examples of exophoric *do this/that* respectively describing an action being performed by the speaker or the addressee. Several of the examples discussed earlier follow this pattern, for instance (260) and (191) (*do this*) or (262) (*do that*). However, there is an asymmetry in the role of this feature, since *do this* may also refer to what the addressee is doing (see (256) or (261) above) whereas *do that* cannot refer to an action being performed by the speaker, as shown in (264). Here, *do this* is preferred, and *do that* would be felicitous if the sentence was uttered after Rodney had finished making breakfast.

5.3. EXOPHORA AND INFERRED ANTECEDENTS

- (264) Rodney cracked two eggs into the frying pan. ‘I’m not doing this/ #that every morning,’ he said. ‘With so many of us we ought to set a rota for cooking.’ (AOR)

5.3.2 Discourse inference

Common to all examples discussed here is that the basis for the interpretation of the VPA is strictly linguistic, but the antecedent is not directly recoverable from a previously-occurring antecedent-trigger, requiring more or less substantial amount of inference from previous discourse.

Some occurrences contain what has sometimes been called a ‘split antecedent’ (Kehler and Ward, 1999; Houser, 2010), in which the antecedent is pieced together from various constituent or nonconstituent sequences. (265) shows a fairly simple case where *do this* is interpreted in terms of the two underlined VPs, with the meaning ‘persecute the Shea Arabs and lay waste to their villages’. (266) is somewhat more complex; the antecedent of *do it*—roughly, ‘present the results of research’—is not as such mentioned in previous context, but it is still inferrable from the two previous passive VPs, which describe how research findings are presented in general and in the author’s own work.

- (265) But er, he [Saddam Hussein] does much, much worse things to his own people, and in particular, the people south, called the Marsh Arabs, or, or the Shea Arabs. He’s continually er, persecuting them, and there’s many, many indications that he uses er, poisoned gas, or chemical weapons, and er lays waste their whole villages, and he’s seeking to do this because he wants to dominate his own country, and not allow any minorities a chance to look after themselves, and the same applies of course to the Kurds in the north, whom we have a direct responsibility, the international community, to protect. (HM4)

(266) Some obvious and important matters were left to one side for years because I couldn't see a way forward, or had no time to do the experiments — or couldn't find the funds to buy the equipment or chemicals needed. Others were picked up opportunistically because a visitor or student arrived with just the right skills or interests to move ahead on a front I might otherwise have neglected. Still others were suggested by a casual reading of someone else's research paper in the train home one evening, or by a talk heard almost by chance at a conference. In some cases an experiment in progress was transformed in design and intention by a result coming from elsewhere in the lab. As Peter Medawar pointed out many years ago in his classic essay *Is the scientific paper a fraud?* [T]hese essential elements in how research is done get refined out from the account as it appears in the finally published papers or scientific reviews, just as they have largely, though not entirely, been filtered from the discussion of Aplysia and LTP in the last chapter. I wasn't proposing to do it that way for anyone's work but my own! (G14)

In other cases, a clear antecedent-trigger is present but not sufficient to interpret the VPA. For instance, in (267), the underlined NP, denotes a state rather than an action, and as such does not directly supply an antecedent for *do this*. It needs to be coerced into an action meaning, roughly 'apply more tightness', so that *do this* can be understood as *tighten the studying nuts*. Examples such as these might be said to exhibit 'semantic mismatch' in addition to the syntactic mismatch between antecedent-trigger and VPA. (268) (a fuller version of 220) shows that such discrepancies in meaning are found with syntactically matched and mismatched cases. In the first occurrence of *do it* the antecedent-trigger is an NP; in the second and last an active VP, while in the third it is a passive VP. Aside from the passive case, where the antecedent is simply 'flood us', *do it* cannot be

5.3. EXOPHORA AND INFERRED ANTECEDENTS

interpreted only in terms of the NP or VP, and requires a causal inference whereby the subject of the VPA clause, in all cases *the gods*, is taken to be responsible for the situation described in the antecedent clause, namely ‘the gods provoked an eclipse/ set fire to the village/created me’. This inference is facilitated by the fact that the antecedent is always introduced in an interrogative asking about the cause of the event (cf. *why*).

- (267) The inductor L2 should not ‘sing’, if it does you may need a fraction more tightness on the studding nuts holding it in place, but do this with extreme care. (C92)
- (268) You have, like it or not, curiosity and with it an adjunct: the desire to explain things. But perhaps these explanations may be a little bit too versatile for modern tastes: Why was there an eclipse?— The gods did it. Why did our village catch fire?— The gods did it. Why have we been flooded?— The gods did it. Why do I exist?— The gods did it. (CET)

Other occurrences require much looser contextual inference to recover the antecedent; unlike in the above examples, the meaning of the VPA is not so closely tied to a specific structure or structures. In (269), the description of the damaged photo frame together with the mention of glue makes salient the action of fixing it referred to by *do it*. In (270), *do this* is interpreted based on the entire first sentence, which describes how the subject tries to keep his balance while travelling on the roof a moving train.

- (269) She lifted out the Salperton photograph. Underneath was a team photograph – a schoolboy cricket eleven with John the second from the right in the back row. As she lifted it out, she realized that the backing was beginning to come away from the heavy cream cardboard of the mount. Perhaps the loft was too damp to store photographs. She would have to mention it to John. She exam-

ined the edge of the mount. All it needed was a little glue. She would do it this evening. (CKB)

- (270) Spreadeagled on the roof of the car in a scissors or St Andrew's Cross position, he concentrated on holding on and not losing his balance as the train gathered speed under the shadow of the Westway and rocked past the desolate terraces of north Kensington. He had done this before, but not here. He had done it on one of the western stretches of the Central Line from North Acton to Ealing Broadway, a rather more hair-raising experience than this. (EDN)

A further class of cases have an explicit description of the action before the VPA, but it occurs much earlier in the discourse, so that the status of that VP (or whatever else the structure is) as the antecedent-trigger is questionable. (271) has the antecedent first introduced in the initial question *How early can science be taught to children?*, but the *do it* sentence occurs only at the end of the addressee's answer. Though it matches the interpretation of the VPA, being so far away, it can hardly be taken as the main or only source of its interpretation. Rather, the antecedent is the central topic of the conversation and is kept salient throughout the utterance by the discussion of what topics might be covered in science class and how they could be taught, both of which are taken to determine the appropriate age for receiving scientific instruction. The VPA clause as a whole also guides comprehension here: the time adjunct *very early* can only make sense in reply to the *how early* question asked by the addressee; it also contrasts with *not very early* earlier in the utterance.

- (271) Mike, how early can science be taught to children?

Well I think that rather depends on saying fairly concisely what it is we mean by science. If in a sense it means how early can you teach children facts and contents and very straightforward knowledge, then I think the answer is not very early at all because it may

5.3. EXOPHORA AND INFERRED ANTECEDENTS

be fairly meaningless that you could teach a child to repeat Newton's law, perhaps the same way as you could teach him to repeat the eleven times table, but without a good concept of number or what Newton meant. It's probably something could learn off parrot fashion, but doesn't have any actual meaning for them. But if you look at science as a way of exploring their world, a world they can structure their curiosity about aspects of the physical world, about aspects of the environment, then I think we can do it very early indeed, probably from the time children can come to school at the age of five and from reception classes onwards. (KRH)

Detailed analysis of the choice between *do it/this/that* in cases of discourse inference would require further investigation, especially as the different situations we have identified are not always clearly distinguishable. *Do it* appears to be the preferred form to refer to a salient action introduced much earlier in discourse, as in (271). Similar usage of *do this* is less typical and seems motivated not by the action being a discourse topic, but rather by the fact that it is taking place at the time of speaking, as with the MPs' planned pay rise in (272). More commonly, as in (270), *do this* derives its meaning from multiple segments in the preceding discourse; in particular, it seems able to refer back to a series of actions that is less salient than those actions taken separately (273).

(272) [Radio phone-in: reactions to proposed pay rise for MPs]

These [MPs] are guys who earn a minimum of thirty one thousand pounds erm if they don't have any companies sponsoring them, a lot of them have companies sponsoring them. They get free travel. Er they don't do so badly at all, they get free er They get allowance for their accommodation in London, Okay? So they're not out of pocket, they don't have to keep two homes going. And I I really don't think they have any business without consulting you and me, as the peoples [*sic.*] that they represent, because I'm I'm a voter

in North Yorkshire too, so I don't think they have any business actually doing this without consulting us first of all. (HUV)

- (273) Basically you will listen for maybe (pause) seven minutes then you switch off for two minutes (pause) then you switch on again and then you try and think well what was I listening to (pause) seven minu- well you know, three minutes ago (pause) what was probably said in the last two minutes. By the time you've sussed that out, you've missed another five minutes, your brain gets confused so you then switch off again for another three minutes and try and clear everything. And you're doing this the whole time. But how do we actually learn? (F88)

As for *do that*, it is not attested in 'topic-maintaining' contexts typical of *do it* and instead behaves similarly to *do this* in (273) with respect to the sort of inference required, see (274) where *do that* means 'break a person's arm with a blow of the wing'.

- (274) 'I remember reading something about geese once,' said Gurder, in a sort of dreamy terror.' It said they could break a human's arm with a blow of their nose.' 'Wing,' said Angalo, looking up at the feathery grey bodies looming over him.' It was their wing,' And it was swans that do that,' said Masklin, weakly. (CEU)

5.4 Conclusion

VPAAs predominantly occur with a syntactically parallel antecedent-trigger (which in the case at hand, is always an active VP, since the corpus contains only active examples) but cases of syntactic mismatch with passive VPs and NPs are also attested. Yet it is even more common for the antecedent to be recovered not from a well-identified linguistic description, but more or less directly from the previous discourse or, less frequently,

from the extralinguistic context.

Statistical evidence so far suggests that VPAs are all equally able to be used with a syntactically mismatched antecedent-trigger; the absence of parallelism does not appear to favour or prevent the use of any particular form. Experimental work would be required to confirm that VPA choice is unaffected by mismatch. However, exophoric reference to an action is most likely to be achieved with *do this/that* rather than *do it*, consistently with [Miller \(2011\)](#)'s suggestion that *do it* requires a higher degree of saliency for the antecedent. Further evidence for this comes from cases where the antecedent is accessed by contextual inference rather than via a specific antecedent-trigger; when this is so, the action referred to by *do it* tends to be highly salient, often by virtue of being a discourse topic. Further investigation into the role of saliency and cognitive accessibility is carried out in the next chapter.

Chapter 6

Saliency and pronoun choice in VP anaphora

6.1 Introduction

This chapter will consider how the choice between *do it* and *do this/that* is affected by the saliency of the antecedent, i.e., how accessible it is assumed to be for the addressee at the point in discourse where the VPA occurs (see [Cornish, 1999](#)). Saliency is itself dependent on the discourse-newness of the antecedent (whether or not it has been mentioned in previous discourse) and its familiarity (whether or not the addressee is thought to have prior knowledge of it). These related concepts have frequently been invoked to account for the choice between various referring expressions ([Prince, 1981](#); [Gundel, 1993](#); [Ariel, 1996](#), amongst others) but they have rarely been studied in relation to VP anaphora (see however [Quirk et al. \(1985, p. 877\)](#), [Huddleston and Pullum \(2002, p. 1534\)](#) and [Miller \(2011\)](#) for preliminary discussion). Under the compositional analysis of VPAs defended in chapter 3, the alternation between *do it/this/that* should be determined by the same properties that determine the choice between *it/this/that* outside of VPA. We will therefore attempt to extend existing

accounts of pronoun choice to VPA alternation. Of particular interest will be the ‘Givenness Hierarchy’ established by Gundel (1993), which makes specific predictions for the use of *it* and *this/that* according to the saliency of the antecedent, namely that *it* requires a more salient antecedent than demonstratives, and further that demonstratives will be infelicitous if the antecedent is sufficiently salient for use of *it*.

We begin by reviewing general properties of *it/this/that* independently of VPA, and consider how they affect the use of *do it/this/that*. We then turn to the question of what makes VPA-type referents more or less salient or familiar, which we will examine in the light of the BNC data on discourse-newness, and consider how Gundel’s model might be extended to account specifically for VPA choice.

6.2 Choice between *it/this/that* outside of VPA

This section focuses on the alternation between *it/this/that* independently of *do it/this/that*. We first discuss the properties of *it* and then review the factors affecting the choice between *this/that* specifically.

6.2.1 Properties of *it*

Various authors consider *it* as well as other third-person pronouns as signalling high saliency for the referent. Biber et al. (1999, p. 331) argue that ‘the interpretation of third-person pronouns frequently requires a good deal of work on the part of the addressee, particularly in conversation’. Based on data from the LSWE (Longman Spoken and Written English) Corpus, they observe a much higher frequency of *it* in conversation than in fiction (respectively 28,000 and 13,000 occurrences per million words). Huddleston and Pullum (2002, p. 1468) view third-person pronouns as definite expressions, explaining that ‘The use of a definite referring expres-

sion implies that the speaker takes the referent to be identifiable without further description’.

Gundel (1993) propose that *it* and *this/that* differ with respect to the degree of saliency they signal for the referent. Using *it* assumes that the referent is ‘in focus’ i.e, in the addressee’s short term memory and at the centre of his/her attention. Pronominal *this/that*, by contrast, are used with referents that are assumed to be ‘activated’—in short-term memory, but not necessarily at the centre of attention. Likewise, in her ‘Accessibility Marking’ Scale, Ariel (1996) ranks unstressed pronouns including *it* higher than demonstratives, meaning that *it* signals a more accessible referent (specifically, *it* is a marker of high accessibility, whereas *this/that* indicate only medium accessibility).

These various accounts tend to associate both *it* and demonstratives with high saliency for the referent, but Gundel and Ariel regard *it* as the preferred form to refer to entities that are especially salient for the addressee.

6.2.2 Choice between *this/that*

The contrast between *this* and *that* has frequently been explained in terms of the relative spatial or temporal distance from the speaker (or sometimes from the addressee, see e.g Biber et al. 1999). Huddleston and Pullum (2002, p. 1504), for instance, argue that the primary use of demonstratives, in their deictic use, is ‘to refer to objects present in the situation of utterance’, with proximal *this* denoting entities close to the speaker, and distal *that* entities further away spatially or temporally. The proximal/distal contrast in spatial location can be illustrated by the following pair from Huddleston and Pullum (2002, 1505, ex. 4i-ii, ; note that they do not use these examples in the context of a discussion of the proximal/distal contrast). (275) is uttered in a context where the addressee is being introduced to Peter and is most likely very close to him. In (276) by contrast,

6.2. CHOICE BETWEEN *IT/THIS/THAT* OUTSIDE OF VPA

over there indicates that the teacher is far from the speaker and addressee, justifying the use of a distal demonstrative. *This* and *that* are clearly not interchangeable here (#*That's my husband/#Isn't this your biology teacher?*).

(275) This is my husband, Peter.

(276) Look over there. Isn't that your biology tutor?

In anaphoric uses, however, this distinction appears to be less central: [Huddleston and Pullum](#) argue that anaphoric *this/that* can generally be replaced by one another 'with little effect on the meaning' (p. 1508). Even in exophoric uses, however, the proximal vs distal contrast is not always sufficient. As [Huddleston and Pullum \(2002\)](#) explain, a speaker holding an object in hand could felicitously ask *What is this?* or *What is that?*. In such a case, they argue, *this* would be the default, and the use of *that* would convey a negative judgement on the part of the speaker. Although negative evaluation may be viewed as a form of distance from the speaker in a figurative sense, it does not change the fact that the spatial location of the referent relative to the speaker is irrelevant to the choice of one or the other demonstrative in this example.

A more central criterion in the choice of *this/that*, at least in their exophoric uses, is whether the speaker's spatial and temporal location is or is not included in the referent: *this week*, but not *that week*, for instance, can refer to the week containing the time of speaking, and similarly *Let's meet next week in this office*, but not *Let's meet next week in that office*, can refer to the office where the conversation is taking place.

[Strauss \(2002\)](#) proposes to account for the use of demonstratives (both determiners and pronouns) and *it* in terms of 'gradient focus', with 'focus' being understood as the amount of attention the addressee should pay to the referent, which is primarily determined by the presumed familiarity of the referent to the addressee as well as its relative importance to the speaker. According to her, *this* signals 'high focus' and denotes referents that are typically new and important information, whereas *that* and *it* sig-

nal medium and low focus respectively, and are used when the referent is more familiar and less important information. Example (277) from [Strauss](#) shows the influence of these parameters in the shift from a distal to a proximal determiner to refer to the same object: in the first two utterances, the discussion centres on Schiavo's experiment, what she intended to do with the bag and what happened to it eventually. The bag is already familiar from having been previously mentioned in the interview, and because both speaker and addressee already know about the experiment. In the last utterance, however, Lauer's question bears on the appearance of the bag, which is much more important to the discussion than in the previous turns. As suggested by [Strauss \(2002\)](#) for other examples, the use of *this* here is also a way to elicit further discussion of the referent, in this case by providing a description of it.

- (277) [Today Show Interview: Matt Lauer and Mary Schiavo, in-studio interview]
 ((Schiavo had conducted an experiment at an Ohio airport in March, 1999, in which she checked a suitcase on a domestic flight and then never boarded the plane; this was to demonstrate safety loopholes on the part of airport security, proving that one could check a bag onto a flight and then never take that flight.))
 Lauer: you expected **that** bag ta go onta Washington withouchu on the plane = and thee airport did (.) stop the bag, didn' they?
 Schiavo: Absolu:tely. Fer whatever reason, **that** bag was sto::pped.
 ((2 skipped turns))
 Lauer: Now were you trying ta make **this- this** package **this** suitcase look like a bomb? (ex. 3, p. 134)¹

1. The transcription scheme used by [Strauss](#) follows the conventions of Conversation Analysis laid out by [Atkinson and Heritage \(1994\)](#), which encode detailed prosodic features such as stress, length or intonation. A partial list of the symbols used is given in the appendix to [Strauss's](#) paper (p. 152), and a more comprehensive one can be found at ac-journal.org/journal/2007/Spring/articles/sensemaking/transcription_symbols.html.

6.2.3 Constraints on the choice between *it* and demonstratives

Gundel (1993)'s analysis of pronoun choice relies on the 'Givenness Hierarchy', a scale on which referring expressions are ranked according to the degree of accessibility they signal for the referent. It is similar in spirit to Ariel's accessibility scale just discussed, but is more specific insofar as it correlates independently-defined levels of accessibility with the use of particular linguistic forms. The Givenness Hierarchy consists of six 'cognitive statuses', which Gundel define as 'assumptions that a cooperative speaker can reasonably make regarding the addressee's knowledge and attention state in the particular context in which the expression is used' (p. 275). Each of the statuses is a necessary condition for felicitous use of the corresponding forms, meaning that use of a particular form signals the referent has the relevant status, and is infelicitous if it does not. They correspond to information about location in memory of the referent and attention state of the addressee—whether or not the referent is in focus of attention, and whether it is present in the addressee's recent or long-term memory. The highest level on the scale, 'in focus', means the referent is present in short-term memory and at the current centre of attention, whereas the lowest, 'type identifiable', means that the addressee can access a representation of the type of the referent. The full set of statuses is shown below in Fig. 6.1:

in focus >	activated >	familiar >	uniquely identifiable >	referential >	type identifiable
<i>it</i>	<i>this/that, this N</i>	<i>that N</i>	<i>the N</i>	<i>indefinite this N</i>	<i>a N</i>

Figure 6.1: The Givenness Hierarchy: Cognitive statuses and associated forms

Here, the relevant forms are in bold, while underlining indicates stress. The equal sign means there is no discernible pause between two words; (.) is for a very short pause, and a colon indicates length. Double parentheses are used to provide contextual information.

Within this model, *it* (as well as unstressed personal pronouns) requires an ‘in focus’ antecedent, while demonstrative pronouns *this/that* require an ‘activated’ referent, i.e, present in short-term memory but not in focus of attention. An illustration of these requirements is provided in the following examples from Gundel:

- (278) My neighbor’s bull mastiff bit a girl on a bike. It’s/That’s the same dog that bit Mary Ben last summer. (ex. 9, p. 280)
- (279) Sears delivered new siding to my neighbors with the bull mastiff. #It’s/ That’s the same dog that bit Mary Ben last summer. (ex. 10, p. 280)

In (278), the dog is introduced as the subject of the first sentence and is likely to also be the topic of the utterance. It can be assumed to be in focus for the addressee, licensing use of *it* (as well as *that*). In (279) by contrast, the referent appears in a PP modifier restricting the reference of the indirect object, and is therefore not the topic of the sentence. It is activated, but not in focus, and can be referred to by *that* but not *it*.

As just mentioned, Gundel argue that each status on the scale is a necessary and sufficient condition for the use of a particular form, meaning that usage of that form is infelicitous if the required status is not met. Additionally, the statuses are implicationally related (i.e, not mutually exclusive), so that an entity which is in focus (of attention) is also activated (in short-term memory), familiar, and so on. Applied to (278), this means that since the referent is salient enough to use *it* (in focus), it is also salient enough to use *that*. More generally, the prediction is that any expression could be used to denote both a referent having the required status (as salient as needed) and one having a higher status (more salient than needed). As Gundel put it, ‘We would thus expect forms to be distributed across more than one status in actual discourse’ (p. 220). For instance, a demonstrative could denote an activated referent as well as an in focus one. Yet if this is so, then the cognitive status of the referent cannot be considered

6.2. CHOICE BETWEEN *IT/THIS/THAT* OUTSIDE OF VPA

a *sufficient* condition for the use of a referring expression. However, in a corpus study, Gundel find that most referring expressions (with the notable exception of definite NPs) primarily or exclusively encode only one status. For instance, just one out of 18 occurrences of *that* in is in focus (the rest are activated), and all 15 occurrences of *this* are activated. Gundel further observe that using a form when the referent is more salient than required is often infelicitous. The following examples from Hegarty (2003) illustrate this problem for the use of *it* vs. *that*. As the GH would predict, *it* can felicitously refer to the snake, which is in focus, but not to the situation of the snake being on the desk, which is only activated, requiring the use of a demonstrative (280a). As (280b) shows, however, *that* is infelicitous in the latter interpretation despite the snake being in focus and thus also activated:

- (280) a. There was a snake on my desk. #It/That scared me.
b. There was a snake on my desk. It/#That scared me.
(ex. 5a-b, p. 895)

These data suggest that other effects restrict the use of referring expressions in discourse beyond cognitive status. Gundel explain their results by appealing to Grice's Maxim of Quantity, which states that the speaker should strive to make his contribution as informative as required (Q1), but not more (Q2). Under this account, using a form when the referent has a higher status than is required would be insufficiently informative since it wrongly implicates that the higher status is not held (in other words, it suggests the referent is less salient than it actually is). In (280b), for instance, using *that* instead of *it* implicates that the referent is activated but not in focus, leading the addressee to search for a less salient antecedent. However, no such effect is observed in (278), where an in focus referent (the dog) can be referred to either by *it* or *that* without leading to an implicature of the sort just described. Gundel argue that implicatures are not necessary inferences, and thus are not expected to arise in all contexts.

This still does not explain precisely why *that* is possible in (278) but not in (280b). This is apparently due to the fact that in (278), unlike (280b) the referent is introduced in a copular structure (*That's the same dog that...*). Outside of such a structure, *that* is impossible, as shown by the following variant:

(281) My neighbor's bull mastiff bit a girl on a bike. #That also bit Mary Ben last summer.

However, it appears that removing the structure in question also makes *that* impossible where the referent is only activated:

(282) Sears delivered new siding to my neighbors with the bull mastiff. #That bit the delivery man.

Essentially, *that* (and also *this*) seems unable to refer anaphorically to first-order entities that are only activated other than in certain types of copular structures. This also applies to inanimates, for instance it is hard to see what *that* refers to in *The book had a strange mark on its cover. #That was green*. Demonstratives can only refer deictically to such entities, for instance one could point at something and say *That just fell*.

6.3 *It/this/that* in VPA

6.3.1 *Do it*: high saliency/old information

There is evidence that *do it* typically recovers antecedents that are highly salient and informationally old. First of all, as predicted by Gundel (1993) for pronominal *it*, *do it* will be infelicitous if the antecedent is not salient enough, and *do this/that* will be preferred. This is often the case in exophoric uses such as (283):

(283) When one of his [a teacher's] blackboards fell onto his foot because Endill had taken the screws from it, he refused to let the class leave

until he had discovered the culprit. ‘Who did this/that#it?’ he shouted. (AMB)

Miller (2011) shows that *do it* is preferred in anaphoric chains to refer back to an antecedent that has already been referred to with a VPA:

(284) I didn’t know I couldn’t do this...so I just did it/#this (ex. 11d, p. edited)

As Miller explains, the unacceptability of *do this* for the second anaphoric mention of the referent is a consequence of the effect described by Gundel, namely that using a demonstrative for a highly salient antecedent in this case wrongly suggests that a less salient one is intended. Further such examples are shown below, with *do that* and (as in the most typical case) *do it* in the antecedent-trigger.

(285) ‘You get all that? It’ll all have to be checked — you’ve got Donalds and Ridley to do that.’ He gazed at her bent head as she read through her notes, and she looked up and smiled faintly at him. ‘Shall I do it now?’ (AB9)

(286) Somewhere quite close a pheasant called. Philip imitated it. The boy stared at him. ‘How d’you do that, then?’ he said. ‘It’s a gift I got,’ said Philip. ‘Do it again.’ Philip did it again. There was an answering call from the far side of the wood. (ABX)

(287) ‘Ryan’s dad smokes, Mom,’ Jo answered at length. ‘Everybody does it.’
‘Your father and I don’t do it, Jocasta, and neither do you.’ (APU)

More generally, it is common for the antecedent of *do it* to be a central topic of the utterance where the VPA occurs, such as in (288), where the antecedent is also situationally salient since Lee has the gun in hand and repeatedly threatens to kill the lamb before asking Philip to do it instead.

At the point where *do it* occurs, the antecedent can therefore be considered ‘in focus’ following Gundel (1993)’s classification.

- (288) ‘What are you doing?’ shouted Philip. The ground was shaking with the galloping of the sheep.
 ‘I’m going to shoot him [a lamb],’ said Lee. ‘Shoot him dead.’
 ‘No,’ said Philip, coming up to him and holding his hand out for the gun.
 ‘I am, I am,’ said Lee.
 And he banged the gun up and down on the gate. Philip looked at where Rebel had separated one of the lambs from the flock. It hadn’t a chance. He said, turning to Philip.
 ‘I can’t see. My glasses are all steamed up.’ He was shaking from the shock of the recoil of the gun, which had sent him flying backwards.
 ‘You do it.’ Philip shook his head. He couldn’t do it.
 ‘Kill it,’ said Lee. (ABX)

Similarly, in (289), the antecedent is again discourse-old and a central topic of the conversation, as shown by the repetition of *kill him*, but it also salient for cultural reasons, since the addressee can be assumed to already know that Rushdie received death threats from Muslim fundamentalists.

- (289) I’m a Muslim fundamentalist, but if I had a knife and I saw Rushdie I wouldn’t kill him, I’d talk to him. A lot of this is political. Plenty of those who talk about killing him would never do it themselves.
 (A1J)

It is of course not always the case that the antecedent is so highly salient and prominent in the discourse. For instance, in (290), the action of washing shorts has not previously been mentioned and is not the topic, but the mention of ‘laundry’ and ‘tights’ in the previous context makes the idea of washing undergarments easily accessible. However, the corpus contains

no examples where the antecedent has especially low saliency as in (283) above.

- (290) I didn't make any enquiries about how the laundry got done until the day I ran out of clean tights. My catering was limited to brewing endless mugs of insipid coffee and opening packets of custard creams. But I got by. In fact, within two years I had gone to the other extreme, washing shorts for lads who were old enough to do it for themselves, and baking cakes for the sole purpose of giving them away. (AHC)

6.3.2 *Do this/that*: speaker-inclusion/exclusion and related features

The criterion of 'speaker-inclusion' discussed earlier for *this/that* appears to be also applicable to *do this/that*. The clearest example of an action that includes the speaker is where he/she is performing that action at the time of speaking. If the action is referred to exophorically, *do that* as well as *do it* are impossible. For instance, if the a speaker is demonstrating an action and invites the addressee to do likewise, only *do this* is possible (291a). On the other hand, if the speaker asks the addressee to imitate him after doing the action, then *do that* is preferred (291b).

- (291) a. [Speaker is juggling with four balls at a time]
Have you ever tried to do this/#that?
- b. [Speaker shows how to juggle with four balls then hands them to addressee]
Now you do that.

(292) is an attested example of *do this* referring to an action the speaker is engaged in.

- (292) Rodney cracked two eggs into the frying pan. 'I'm not doing this/

CHAPTER 6. SALIENCY AND PRONOUN CHOICE IN VP ANAPHORA

#that/#it every morning' he said. 'With so many of us we ought to set a rota for cooking.' (AOR)

The speaker may also be considered as included in the event if he is not the agent performing the action, but a participant in the event, e.g., in (293) the patient and in (294) the beneficiary of the offer of accommodation.

(293) — What were they doing to you?

— Er I was handcuffed to a radiator in the nude and urinated on. I was tied into a sleeping bag and hung upside down from a tree overnight. Er

— Why did they pick on you do you think Andrew, er James?

— My size.

— Yeah.

— And you know

— Did you s— did you te— did you say for God's sake stop doing this to me? Did you try and stop them, did you try and make them see reason? (HVD)

(294) Look, I've got a spare room. You can use it if you want.'

'Thanks, but...' she trailed off with an awkward shrug.

'I mean, I don't even know you.'

'Likewise,' Bernard replied. He bit his lip thoughtfully.

'I'll tell you what. Call your friends and see if they can put you up for the night. If they can't you can either stay at the flat or else I'll give you some money and drop you off at a hotel.'

'Why are you doing this?'

'My father raised me. I never knew my mother. He was the only family I had.' (EF1)

The action does not necessarily occur at the exact time and place of speaking; it can simply contain them. In (295), for instance, the speaker

is referring to current political events in Britain.

- (295) ‘Damn Gladstone! Damn the Liberals!’ His voice shook with anger.’ I knew what was going to happen more than twenty years ago. I could see all this coming. And now this damned Wyndham Act... God! That I should live to see the day when a British government would do this to us. Don’t they know – don’t they realise? Allowing the Irish peasants to buy land – did the government think they would be satisfied with just that? (B1X)

In third-person narratives, *do this* is often used to achieve focalisation through the point of view of the agent (296), or in some cases of an internal narrator.

- (296) Amanda inspected her drink. To do this she unwrapped her legs from each other, bent her right knee briefly and peered into the empty glass. (AOR)
- (297) His concentration was total. He moved along the row, putting in money, pulling the handle, watching the drums revolve and click into place, moving on. When he reached the last machine he went back to the beginning and played them all again. As they watched, he did this over and over. (ACB)

Conversely, *do that* tends to denote actions that exclude the speaker in various ways. For instance, it may be used to refer to the addressee’s actions (298), or to describe what the speaker is unable, unlikely or unwilling to do, and so on (299), or to actions regarded as such more generally, not just for the speaker. (300) shows a combination of these features, as *do that* denotes an action which is possible for the addressee but not the speaker.

- (298) Somewhere quite close a pheasant called. Philip imitated it. The boy stared at him. ‘How d’you do that, then?’ he said. ‘It’s a gift

I got,' said Philip. 'Do it again.' Philip did it again. There was an answering call from the far side of the wood. (ABX)

- (299) Early on in that [Hugo's *Les Misérables*], the hero's homeless and somebody puts him up for the night.' She looked me straight in the eye. 'I don't have the courage to do that, Dorothy.' There was a pause. I smiled. 'Neither would I,' I said.' (A0F)
- (300) Don't knock yourself. I mean, you use words like 'incandescent' and 'eminently'. I wish I could do that. I'd really like to be able to speak like you, you know. You may be homeless but at least you're not a dumb blonde like I am.'(A0F)

Temporal remoteness of the action, which would be an example of the distal meaning of *that*, may also be considered as excluding the speaker insofar as it excludes the time of speaking:

- (301) She gazed into her glass. 'I'm going to have another drink. It's so nice in here, not worrying about other people, or getting supper, or anything. And sitting with you. It's a long time since we did that.'

6.3.3 Topic continuation and importance of the referent

The use of *do this/that* appears to be explainable at least in some cases by the relative importance of the antecedent and whether or not it is continued as a discourse topic. *Do this* is commonly followed by further discussion of the antecedent, in some cases quite extensive, thus framing it as important to the conversation. (302) is an example where the antecedent is important because it is a recent scientific finding, thus new knowledge, and has also led to another discovery, warranting further discussion following the VPA. *Do that* remains possible here, but would not similarly convey that the referent is important and will be discussed further.

(302) There is, indeed, a snake — the Formosan banded krait — that gives you a type of instant miocenia [myasthenia], and the discovery that the venom of this animal contains a toxin that can do this has, indirectly, led to the elucidation of the mechanism behind *miocencia gravis*. (KRF)

By contrast, *do that* tends to close off discussion of the antecedent and may contribute to changing the topic, such as in (303) where the sentence following the *do that* sentence does not further discuss the activity of watching voles, but instead gives a description of the animal.

(303) Even 20 years ago, Rob Strachan recalls, the vole was much more common. There were rivers he visited as a child where he could sit and watch water voles every day. He cannot do that now. The water vole is about the same size as a brown rat, but with blunter features...(AJK)

6.4 On the saliency of actions and events

Earlier in this chapter we defined the saliency of the antecedent as a function of the ease with which the addressee is expected to retrieve it at the point in discourse where the VPA occurs. The question is then what makes an antecedent more or less salient in the sense understood here. We will consider both general factors affecting the saliency of a referent, as discussed by Gundel (1993) and others, as well as those applying more specifically to actions and events. We first investigate how the linguistic context and the speech situation can contribute to making the antecedent salient, and then discuss the role of the addressee's familiarity with the referent.

6.4.1 Linguistic factors

As suggested by [Borthen et al. \(1997\)](#), a referent that has already been mentioned is more salient than one which is new to the discourse. It can be further argued that the saliency of the antecedent increases with each subsequent mention, keeping it in the addressee's short-term memory and possibly in focus of attention. On the other hand, the absence of previous mentions of the antecedent does not automatically render it less accessible. The same is observed with VPAs: *do it* appears to be favoured if the antecedent has already been mentioned before the antecedent clause, but is not restricted to such a context.

Topicality is another way in which the antecedent can become more salient. According to [Gundel \(1993\)](#), a referent is more likely to be in focus of attention if it is the topic of the utterance and is likely to be continued as a topic of subsequent utterances. Though it is not common for VP referents to become topics, multiple references to the antecedent can contribute to establishing and subsequently maintaining it as a discourse topic, as was observed in (288) above.

[Gundel \(1993\)](#) as well as [Hegarty \(2003\)](#) argue that syntactic structure also influences the saliency of a referent. For instance, as discussed with (278)–(279) above, the referent of a subject NP is more salient than one occurring in a less prominent syntactic position, such as inside a modifier. The influence of syntax on the saliency of VP remains unclear, especially as the analysis in [Gundel \(1993\)](#) is limited to NP referring expressions, mostly when they denote first-order entities. [Hegarty \(2003\)](#), following [Gundel et al. \(1999\)](#) shows that a clausally-introduced event is sufficiently salient upon its first mention in the discourse to be referred to by *it* as well as a demonstrative, as shown in (304):

(304) John broke a priceless vase. It/That happened at noon. (adapted from [Hegarty \(2003, ex. 13, p. 898\)](#))

6.4. ON THE SALIENCY OF ACTIONS AND EVENTS

This appears to be also true of actions introduced in the main VP of the sentence (305):

- (305) John refurbished his living-room. It/this/that took him a lot of time.

Detailed investigation of the way in which syntactic structure affects the saliency of VPA antecedents is beyond the scope of this thesis. It may be noted, however, that just as described by Gundel (1993) for NPs, an antecedent introduced in a lower syntactic position is less salient, as shown by the impossibility of *do it* in (306):

- (306) They [gang members posing as Japanese businessmen] were losing about 10 per cent of their profit by not claiming back the tax, and no Japanese businessman does that/this/#it. (K4V)

6.4.2 Extralinguistic factors

The primary way in which an action can be made salient by the extralinguistic context is if it is ongoing at the time and place of speaking. However, this does not give it high saliency, as in most cases, only *do this/that* can refer exophorically to such actions:

- (307) ‘What you doin’ here?’ ‘We’re dyeing popcorn,’ Toby said grandly. John was fascinated. He moved between us and looked at the bowls of deep colour, and little puffs of red and yellow dotted about the draining-board to dry.’ ‘But what you do this for?’ ‘We’re going to string them together and hang them on our Christmas tree.’ (FEE)
- (308) Herr Nordern closed the door. As he did so his wife raised her eyebrows as if at some freakish aberration. ‘What on earth are you doing that/#it for?’ she asked. (A7A)

Do it only seems possible if the addressee is watching the action as it occurs, consistently with Gundel's in focus requirement (309). Note that here, the antecedent is made accessible by a combination of the extralinguistic situation and the mention of 'edging it [= the sofa bed] round the corner', but *do it* remains possible even without the previous two utterances, as shown in the constructed variant (310):

- (309) [Ludens and others are trying to move a sofa bed out of a room]
 'There must be some method of edging it round the corner,' said Ludens.
 'You keep saying so, why can't you find it!'
 'They don't know how to do it,' said Patrick conversationally to Franca. They were standing on the lower landing watching. (APM)
- (310) [Patrick and Franca are watching Ludens and others trying to to move a sofa bed out of the room]
 'They don't know how to do it,' said Patrick.

An action may also be salient if it has just occurred or is expected to occur soon. The corpus only contains examples of the first type, such as (311)–(312). As with ongoing actions, *do it* is infelicitous if the event is not at the centre of attention.

- (311) Frau Nordern turned her face away. 'Don't do that/#it, Helga,' Herr Nordern said. 'Look at me, please. Thank you.' (A7A)
- (312) Somewhere quite close a pheasant called. Philip imitated it. The boy stared at him. 'How d'you do that/#it, then?' he said. (ABX)

Actions occurring at a very different time from the time of speaking can be assumed to have relatively low accessibility, as they may not even be in the addressee's short-term memory. In some cases they can be inferred from the speech situation if they are made salient by features of the extralinguistic context such as in (313) or (314). In the absence of such

6.4. ON THE SALIENCY OF ACTIONS AND EVENTS

clues, reference with any VPA seems impossible; e.g., *do this* would not be felicitous if the class had not seen the blackboard fall in (313), and in (314) it is felicitous because Sam and the addressee are both inspecting the damage to the door.

(313) When one of his [a teacher's] blackboards fell onto his foot because Endill had taken the screws from it, he refused to let the class leave until he had discovered the culprit. 'Who did this?' he shouted. 'Tell me now or you'll all have three lengths of the corridor.' Everyone looked at each other, but no-one spoke.

(314) Sam disgustedly fingered the splintered door frame. 'Did you sodding do this?' he demanded.

'It wasn't locked.'

'It was,' I said. 'With no key in sight.'

'The key was in the keyhole on the inside.' (ADY)

6.4.3 Familiarity beyond discourse

Assuming that the addressee is familiar with the antecedent means assuming that he has previous knowledge of it independently of the current exchange. What matters here is familiarity with the event or situation, rather than the action; for instance, in the following, saying that the antecedent of the two *do it* occurrences is familiar means that the addressee is aware that many people smoke, not just that she is familiar with the action of smoking itself.

(315) 'There were people taking drugs at that party, I could smell it.'
'That was patchouli oil, Mom, can't you tell the difference? God, you're so naive, I can't believe it. Anyway, so what if there were people smoking a little—everybody does it, the teachers in school do it...'

More often than not, it is difficult to determine whether or not the addressee is familiar with the antecedent, and which assumptions are being made by the speaker in this respect, unless such assumptions are clearly stated by the speaker. In some cases, the antecedent can be presumed familiar or unfamiliar for cultural reasons. In (316), the writer reports on a rural custom most likely unknown to the reader, whereas in (317), as discussed, the addressee can be assumed to already know that Rushdie received death threats from fundamentalists.

- (316) Once a year in this village men open the door to let the ghosts in. There is nothing unique about this; men are compelled by law to do it in almost every English village, and it is not a matter of one ghost or two, but a whole rout of them — quiet chaps in cowls, roaring boys in ruffs, ladies in farthingales.
- (317) I'm a Muslim fundamentalist, but if I had a knife and I saw Rushdie I wouldn't kill him, I'd talk to him. A lot of this is political. Plenty of those who talk about killing him would never do it themselves.
(A1J)

To a certain extent, assumptions about the familiarity of the antecedent may also be encoded linguistically through the use of certain structures signalling a particular informational status for the referent; for instance, the VP[to] predicate of clefts, pseudo-cleft or similar constructions is usually new or contrastive, the first part (open proposition) being given:

- (318) What I'm trying to say is that it's providing for people already here that I think should be a major item of population policy, and doing it not in a massive world sense, or even a country, but in particular areas as well. (KRE)
- (319) So what you do is, to type check disk space slash and then it [the program] asks you wh- if it comes up with a question, which I expect it may well, if you've never done this before. (HDV)

Though it may be argued that familiarity can increase the saliency of the antecedent to some extent, it is not as strong a factor as previous mention of the antecedent in discourse, and even a completely unfamiliar referent may be highly salient.

6.5 Discourse status of the antecedent

Discourse-newness of the antecedent was assessed at the level of the sentence containing the antecedent-trigger, based on the presence or absence of previous reference to the antecedent in earlier discourse. For each occurrence, the antecedent was classified as discourse-old if it had been mentioned at least once in the preceding context, and discourse-new otherwise. In exophoric uses, the VPA constitutes the first mention of the antecedent, which is therefore new to the discourse. Discourse-old cases also include those where the antecedent is introduced by a pronoun (320) or an elliptical element such a null complement in NCA (321), or ellipted material following an auxiliary in VPE (322):

- (320) For instance we know that, that erm in this country you can't get a mortgage without taking out fire insurance, and the reason why it's mandatory in a mortgage is because if it were left to people's whim they wouldn't do it, they'd say, 'Well it won't happen to me,' and mortgage companies don't like that very much. (KRH)
- (321) 'Do you want to open the wine?' Kate nodded, and rummaged in the carrier bag for the bottle and the corkscrew. 'Can you manage ∅? I'll do it if you want.' (CKB)
- (322) How could he keep Andrew at Agricultural college? Well they would, somehow or other. They'd done it for Adam and Christopher and they'd do it for him. (AC2)

In a small minority of occurrences, the discourse status of the an-

tecedent could not be determined and was marked ‘NA’. This was the case when the VPA occurs in a portion of quoted direct speech that does not include the antecedent-trigger, as in (323), (even if a linguistic description of the antecedent is provided outside of the quoted segment), or conversely, in fiction texts, if a VPA in a passage of free indirect speech refers back to an action introduced in dialogue.

(323) [BROADCASTER] Tonight, business angels, the private investors who’ve made millions by backing risk and innovation in partnership with Britain’s small entrepreneurs.

[BUSINESS ANGEL] I think I do it because (pause) I’m so depressed by the poor performance of professional investment managers in the city, and I hope I can do better than they can. (HM3)

(324) They knew when they were opposite Pitcaple, and then Cluny, by the barking of the dogs and the rattle of their chains. ‘They would have our throats out if they could get at us.’ Allan Stewart’s voice sounded tiredly out of the darkness. Cameron had not the energy to reply, but he was thinking, So would their masters. Dogs would do it to us with their teeth but they have laws and guns instead. (A0N)

Table 6.1 below, reproduced from chapter 2, shows the distribution of discourse-old and discourse-new antecedents for each VPA along with cases where the discourse status is unclear. It is immediately apparent that *do this/that* typically have discourse-new antecedents, whereas *do it* shows a weaker preference for discourse-old antecedents. All reported preferences were found to be significant (*do it*: $p < 0.001$, *do this/that* $p < 0.001$). The difference between *do this* and *do that* was also significant, meaning that a discourse-new antecedent is especially typical of *do this*. The interaction between discourse status and VPA choice was significant ($p < 0.001$).

In short, while *do this/that* appear closely associated with the signalling

6.5. DISCOURSE STATUS OF THE ANTECEDENT

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	Total
New	196 (39.2%)	160 (80%)	147 (73.5%)	503 (55.9%)
Old	292 (58.4%)	31 (15.5%)	53 (26.5%)	376 (41.8%)
NA	12 (2.4%)	9 (4.5%)	0	21 (2.3%)
Total	500	200	200	900

Table 6.1: Discourse status of antecedent by VPA

of new rather than old information, *do it* seems equally liable to express both, although it is likely to be preferred over *do this/that* if the antecedent is old. This is consistent with the fact that a discourse-old antecedent is also more accessible, and thus more likely to be referred to with *it*. The data also lend support to [Strauss's](#) proposed contrast between *this* and *that*, with *do that* being in between *do it* and *do this* as regards the discourse status of the antecedent.

Discourse-newness varies considerably between the written and spoken sections of the corpus, as [Table 6.2](#) below indicates. Spoken texts show a higher ratio of discourse-old antecedents, whereas written texts have a much higher ratio of discourse-new antecedents. The interaction between text type and discourse newness was significant ($p < 0.0001$).

	Old	New	NA	Total
Spoken	241 (53.6%)	194 (43.1%)	15 (3.3%)	450
Written	135 (30%)	309 (68.7%)	6 (1.3%)	450
TOTAL	376 (41.8%)	503 (55.9%)	21 (2.3%)	900

Table 6.2: Discourse-newness of VPA antecedents by register

In the data from the written sample only ([6.3](#)), both dialogue and narration show a higher ratio of discourse-new antecedents, but this preference is stronger in narration. The interaction between register and discourse newness in the written BNC was significant ($p < 0.001$).

By comparison, data for the entire BNC reported in [Table 6.4](#) (in which

	Old	New	NA	Total
Dialogue	88 (38.8%)	136 (59.9%)	3 (1.3%)	227
Narration	47 (21.1%)	173 (77.6%)	3 (1.3%)	223
TOTAL	135 (30%)	309 (68.7%)	6 (1.3%)	450

Table 6.3: Discourse-newness of antecedent in dialogue and narration (written sample)

‘Dialogue’ comprises both fictional dialogues and actual spoken data) also show a higher ratio of discourse-old antecedents in dialogue than narration, but no preference between discourse old/new in dialogue. The interaction between register and discourse-newness was also significant here ($p < 0.0001$).

	Old	New	NA	Total
Dialogue	330 (48.7%)	329 (48.6%)	18 (2.7%)	677
Narration	47 (21.1%)	173 (77.6%)	3 (1.3%)	223
TOTAL	376 (41.8%)	503 (55.8%)	21 (2.3%)	900

Table 6.4: Discourse-newness of antecedents in dialogue and narration (overall)

The observed discrepancies between registers may be explained by the fact that spoken discourse tends to be more repetitive and only infrequently introduce new topics, whereas written discourse, and fiction especially, tends to avoid such repetition.

6.5.1 Discourse-old antecedents with *do this* and *do that*

Given the preference of *do this/that* for discourse-new antecedents, their use with discourse-old antecedents may be regarded as exceptional to some extent, especially so for *do this*. Such uses appear to be largely explainable in terms of the discourse properties of demonstratives discussed earlier which may lead to a preference for *do this/that* over *do it*.

6.5.1.1 *Do this*

The use of *do this* with a discourse-old antecedent is extremely rare in fiction—just three occurrences have an antecedent that is already mentioned prior to the antecedent-trigger clause. In two of them, such as (325), the use of *do this* seems driven by the choice of narrative point of view, with the action being viewed from the subject’s perspective rather than the narrator’s. Replacing *do this* with *do that* here would imply that the whole *do this* sentence is the narrator’s assertion, rather than the character’s own assessment of the situation.

(325) When she looked down into the channel with Gazzer and saw Simon, water swilling round his waist, she could hardly believe that she was responsible. It seemed as if another person had done this, not her. ‘You must’ve been crazy!’ she told herself. (ACB)

The third occurrence has *do this* in a passage of quoted speech; though the antecedent-trigger is unavailable, it can be assumed beyond doubt that the antecedent has already been discussed. Both *do it* as well as *do that* are possible, but *do this* highlights the fact that the action occurred recently and is important to the speaker.

(326) I’ve just been over to see Madge and guess what? Her house was burgled last night! Really turned over. They’d nicked money—smashed things up. Mr Bishop’s brother told me. He said it made you sick to look at it. He’d been round there early this morning to get some papers for the solicitor or summat. Anyway, do you know what he said ? He said: ‘Thank God Madge wasn’t there. Whoever did this was a right nutter.’ (ACB)

This also appears to be the main reason for the use of *do this* with discourse-old antecedents in actual dialogues, where it tends to signal or elicit further discussion of the antecedent (327). Repetition of *do this*, in

particular, may be seen as a way of ‘pressing the issue’ to maintain the antecedent as a discourse topic (328).

- (327) The last sheaf [of corn] was cut then it was er put in the barn and kept there for the year and that was called the maiden sheaf.
 — Mm. And what was the purpose behind the custom?
 — Oh it was just er just I suppose luck, just an idea to have the last sheaf it was called the maiden sheaf.
 — Mhm. And every farmer up the glen would do this?
 — Oh yes everybody had their last sheaf. (G62)
- (328) Well I think a lot of evaluation does go on [*in education*]. I mean I think part of being a proper professional means that one does attempt to evaluate what one is doing, and I think that most teachers do do this. I think that the issue really is erm are there ways in which perhaps they could be helped to do this more productively, and are there ways in which they could be helped to do this rather more collaboratively than perhaps they have done so in the past?
 (KRH)

6.5.1.2 *Do that*

Use of *do that* with discourse-old antecedents is in some cases linked to the presence of a structure preventing *do it*, typically an adjunct-less *if*-clauses, or by the low saliency of the antecedent:

- (329) Sometimes Marie writes some stuff—poems and that, or else she reads a book. She likes reading. If she does that, I get on with the jigsaw. (A74)
- (330) So what’s what’s your occupation Martin, are you a songwriter?
 Erm well not as such as a songwriter, I mean I do that yes but I’m my occupation I’m a self-employed window cleaner. (HMD)

6.5. DISCOURSE STATUS OF THE ANTECEDENT

Most of the time, however, the relevant feature seems to be speaker-exclusion or distality, as proposed earlier for *that* as well as *do that*, namely actions in which the speaker is not involved, or which he is unlikely or unwilling to do, and so on.

(331) ‘And then,’ said the boy, his eyes on Doyle’s face, ‘then you’ll shoot me. Whoever I am.’

‘Oh, I shan’t shoot you,’ Doyle murmured lightly. ‘Your mother will do that.’ Jinny did not understand, but the boy’s head turned and he stared at the Hare-woman. ‘Ma? You would really –?’ (AC4)

(332) Unfortunately, PC Dickens happened to be a writer too. Although he was no relation to Charles, his surname had inspired him to pick up the pen. He insisted on sending some of his ‘character sketches’ about life in the Force to Nigel. He didn’t need help placing them — the police have so many interesting in-house magazines with pictures of stolen objets d’art and jokes and stories in the back. All he wanted, he said, was a little constructive criticism. Nigel went back home empty-handed and downhearted. He really hated literary wankers who sent him stuff to criticize. He’d never done that to anybody. What had he done in his last incarnation to deserve it? (AC3)

Less commonly, *do that* emphasises the temporal remoteness of the antecedent action, as made clear by *that was a long time ago* in (333).

(333) — If David Owen is talking to terrorists, why don’t the British government talk to Gerry Addams?
— They did and look where it got them.
— They did that in 1972, that’s right
— In 1972. That was a long time ago, there’s been a lot of water under the bridge, why not talk again? (HV2)

6.5.2 Conclusion: discourse-newness and VPAs

The discourse status of the antecedent (previously mentioned or not before the antecedent clause) plays an important role in the choice between VPAs. *Do this/that* most often refer back to antecedents that have not been previously mentioned, whereas *do it* tends to be preferred if there is previous mention of the antecedent. The influence of discourse status is especially strong with *do this/that*, which rarely occur with already-mentioned antecedents, whereas *do it* is favoured by, but not mostly limited to, contexts where the antecedent has already been mentioned. When *do this/that* occur in such contexts, they can usually be replaced with *do it*, and the use of a demonstrative is motivated by the need to achieve certain discourse effects. *Do this* can signal the antecedent as important and maintain it as a topic, whereas *do that* may contribute to closing off discussion of the antecedent and possibly changing the topic, or, more commonly, present the action as unlikely, striking, objectionable or otherwise far from the speaker.

6.6 Saliency

The previous section investigated how the choice of *do it* or *do this/that* is affected by whether the antecedent has already been mentioned before the antecedent clause. Here, we examine the role played by the saliency of the antecedent, understood as its presumed accessibility for the addressee at the point where the anaphor occurs. Specifically, we will consider how the conditions on the use of *it/this/that* proposed by Gundel (1993) based on their Givenness Hierarchy (henceforth GH) can be extended to the alternation between VPA *do it/this/that*.

6.6.1 Corpus study: applying the GH to VP anaphors

6.6.1.1 Corpus design and annotation scheme

To assess the usefulness of the GH in accounting for VPA alternation, a corpus study was conducted on a sample of data from the BNC in which the cognitive status of the antecedent was annotated for each occurrence. Since the annotation was carried out by the author, it was essential to avoid using occurrences from the initial corpus of 900 examples, as these were already very familiar, which might have interfered with the coding. Instead, a sample of 120 occurrences—40 each of *do it*, *do this* and *do that*—was collected from the Newspaper section of the BNC. Aside from being separate from the initial set of data, the examples from press articles also have the advantage of requiring less extensive context to be understood, contrary to the fiction and spoken examples in the larger corpus.

The coding of cognitive statuses relied on the definitions provided by [Gundel](#) as well as their comments on the examples in their paper. To avoid determining the cognitive status based on the actual pronoun occurring in the example (for instance, coding the antecedent as ‘in focus’ because *do it* is used), the VPA was systematically replaced with *do X* to obscure the object pronoun², and the data were randomly sorted.

Following the discussion in [6.2.3](#), we would expect *do it* to occur primarily when the antecedent is in focus, and *do this/that* when it is activated. Since an activated antecedent would be insufficiently salient for *do it*, there should be very few attested examples. While *do this/that* may in theory be used with an in focus attested cases should also be rare since the use of a demonstrative may wrongly suggest the antecedent is only activated and not in focus.

2. While this precaution is helpful, it does not completely remove bias, since the coding may still be influenced by the pronoun which the coder thinks is likely to be used in that particular context.

6.6.1.2 Results and discussion

Table 6.5 shows the distribution of cognitive statuses across VPAs. Only activated and in focus were used; we found no occurrences whose antecedent had the status ‘familiar’ (in long-term memory, but not necessarily activated) or below.³ In focus is the most frequent status overall (57.5% of tokens) and with all forms except *do that*. Statistically, a significant interaction between cognitive status and VPA choice was observed ($p = 0.005$). The preference for in focus was also significant with *do it* ($p < 0.001$), but not *do this* ($p = 0.75$). The preference for activated status with *do that* was not significant ($p = 0.34$). Pairwise comparisons between VPAs showed a significant difference between *do it* and each of the demonstrative variants (*do it/this*: $p = 0.03$, *do it/that*; $p = 0.03$), but not between *do this/that* themselves ($p = 0.5$).

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
In focus	31 (77.5%)	21 (52.5%)	17 (42.5%)	69 (57.5%)
Activated	9 (22.5%)	19 (47.5%)	23 (57.5%)	51 (42.5%)
Total	40	40	40	120

Table 6.5: Cognitive statuses by VPA

These results only partly conform to the predictions of the GH. As expected, *do it* mainly occurs with highly salient antecedents, whereas *do this/that* are more common if the antecedent is less salient. However, contrary to what the GH would predict, *do this/that* are also commonly used with in focus antecedents without implicating that a less salient referent is intended, suggesting that their primary function is not to signal low saliency (or specifically that the antecedent is not in focus of attention).

3. The choice of a newspaper corpus makes exophoric examples less likely to occur than in a spoken corpus. Exophoric examples are not discussed by Gundel et al., but they appear to have ‘familiar’ status. It may be the case that further examination of spoken data would provide attested cases of this type and that their absence in the present corpus is due to register and size.

Where the antecedent of *do it* is activated, it is usually recently mentioned in the discourse, making it more accessible even though it is not sufficiently salient to achieve in focus status. In (334), for instance, the antecedent ‘hang the wrong person’ is unlikely to be in focus of attention, but it is related to the more general topic of hanging. Similarly, ‘washing shorts’ in (335) has only been mentioned once and is therefore only activated, but it is an example of doing the laundry, which is the main topic of the utterance.

(334) I am convinced that if hanging were reintroduced, they would not just occasionally hang the wrong person, but do it pretty well every time. (AHX)

(335) I didn’t make any enquiries about how the laundry got done until the day I ran out of clean tights. My catering was limited to brewing endless mugs of insipid coffee and opening packets of custard creams. But I got by. In fact, within two years I had gone to the other extreme, washing shorts for lads who were old enough to do it for themselves, and baking cakes for the sole purpose of giving them away. (AHC)

Outside of such cases, *do it* tends to be infelicitous if the antecedent is not in focus, such as in (336) where the antecedent is in a means adjunct, making it less accessible despite its textual proximity:

(336) They [gang members posing as Japanese businessmen] were losing about 10 per cent of their profit by not claiming back the tax, and no Japanese businessman does that/this/#it. (K4V)

Regarding the use of *do this/that* with in focus antecedents, the following examples suggest that *do it* is generally an acceptable alternative, though it is unclear at this point what motivates the use of *do this/that* in the first place:

(337) When Mr Gummer took over the Ministry of Agriculture, he promised

he would put the consumer first. His chance to do this(/it) has come with the microwave report, but he has decided to put the interests of microwave manufacturers first. (AAX)

- (338) We will introduce the policies that we have undertaken to introduce, and in the course of doing that(/it) I hope we can attract a breadth of support to add to our majority and to add to the degree of consensus supporting that majority (...) (AJ6)

On the other hand, replacing in focus *do it* with *do this/that* may or may not be acceptable. While it might be assumed that they are dispreferred when the antecedent is highly in focus (for instance when it is mentioned several times before or after the anaphoric clause, becoming a central topic of the utterance), this is not necessarily the case:

- (339) I'm a Muslim fundamentalist, but if I had a knife and I saw Rushdie I wouldn't kill him, I'd talk to him. A lot of this is political. Plenty of those who talk about killing him would never do it#this/#that themselves. (A1J)
- (340) Surely we could return Brightness the Beluga to her own genetic family in the Arctic? We, the British, had the expertise. We could mobilise the resources and we had the moral drive. So plans were laid to do it/this/that. (AJS)

More strikingly, there are some cases of in focus *do this/that* where *do it* is infelicitous. Although such examples are rare, they are clearly in contradiction with the GH, as it is not expected that a form will be dispreferred despite the required status being met:

- (341) While Christians in Britain lapse, in Germany (...) they formally notify the local authorities of their Austritt (resignation). Every baptised German who has not done this/#it and who pays income tax must also pay the additional church tax. (AK9)

(342) He closes his eyes when he speaks and I don't trust anyone who does that/#it. (AHF)

In short, the results of the cognitive status coding suggest that saliency as conceived of in the Givenness Hierarchy plays only a limited role in VPA alternation. Though *do it* behaves as expected and occurs mostly with in focus antecedents, it also sometimes allow activated ones. *Do this* and *do that* are more typical if the antecedent is activated, but they readily allow in focus antecedents as well, and may sometimes be substituted for *do it* if the antecedent is in focus. Examples like (341)–(342) above are left unexplained by the GH, since the cognitive status of the antecedent is assumed to be at least a sufficient condition for the use of a pronoun.

6.6.2 Alternative coding: saliency levels

Since the initial annotation according to cognitive statuses provided mixed results, a second coding of the same data using different criteria was carried out by Ana Perlstein, research assistant at Paris Diderot University. The format of the data was the same as in the first coding (randomly-sorted examples with *do X* instead of *do it/this/that*). The second coder did not have access to the initial corpus and was therefore unaware of the number of occurrences of each VPA. The annotation did not make use of a specific framework, but simply reflected the coder's intuition of how accessible the antecedent was, with a three-way distinction between 'most salient antecedent' (MSA), 'highly salient antecedent' (HSA) and 'activated antecedent' (AA).

Table 6.6 shows the distribution of saliency levels across VPAs. In the vast majority of cases, the antecedent was coded 'most salient'; 'highly salient' was used only rarely and 'activated' for just three occurrences. Unlike in the cognitive status coding, the frequency of the highest level is virtually the same for all VPAs; *do this/that* do not appear to be favoured by

less salient antecedents. The interaction between VPA choice and saliency was not significant ($p = 0.59$), and no significant differences between VPAs were observed (*do it/this*: $p = 0.6$, *do it/that*: $p = 0.3$, *do this/that*: $p = 1$).

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	Total
Most salient	30 (75%)	32 (80%)	31 (77.5%)	93 (77.5%)
Highly salient	10 (25%)	7 (17.5%)	7 (17.5%)	24 (20%)
Accessible	0	1 (2.5%)	2 (5%)	3 (2.5%)
TOTAL	40	40	40	120

Table 6.6: Saliency levels by VPA

To a certain extent, the results of the alternative coding contradict those of the first, which showed a significant but limited effect of givenness. However, the main criterion distinguishing activated from in focus in the GH is whether the referent is at the centre of the addressee’s attention, whereas here it is primarily the degree of accessibility which matters. The next section compares the two annotation schemes in greater detail.

6.6.3 Cognitive statuses and saliency

Whereas the initial coding used just two of Gundel’s statuses, the alternative coding had three levels, raising the question of how these levels map onto cognitive statuses. ‘Most salient’ and ‘activated’ are comparable to ‘in focus’ and ‘activated’ respectively, while ‘highly salient’ may be viewed as in intermediate status between ‘in focus’ and ‘activated’.

Table 6.7 gives a more precise overview of the relation between cognitive statuses and saliency levels. ‘Most salient’ was used primarily with in focus occurrences, but also in a number of the activated cases. The difference was close to significance ($p = 0.07$). The ‘highly salient’ level was also commonly applied to both in focus and activated cases, though the preference for in focus status is less strong and not significant ($p = 0.7$).

Interestingly, the ‘activated’ level was applied once to an antecedent initially coded as in focus. In other words, there is a relative degree of overlap between ‘highly/most salient’ and ‘in focus’, but the conception of what constitutes an activated antecedent differs between the two coders.

	In focus	Activated	Total
MSA	55 (59.1%)	38 (40.9%)	93
HSA	13 (54.2%)	11 (45.8%)	24
AA	1 (33.3%)	2 (66.7%)	3
Total	69 (57.5%)	51 (42.5%)	120

Table 6.7: Cognitive status by saliency level

Closer examination of the data reveals that recent mention of the antecedent was a central criterion used by the second coder for determining saliency in the alternative. In the vast majority of cases the second coder classified as ‘most salient’, the VPA occurs either in the same sentence as the antecedent-trigger (343), or (more commonly) in the next sentence (344). In the latter case, the following sentence tends to be very short (345):

- (343) This chemical is usually turned into an odourless substance, but some unfortunate carriers of a faulty gene cannot do this – and the unaltered chemical is secreted in breath and sweat, causing the sufferer to smell strongly of fish. (K2V)
- (344) This talented orchestra can only survive if the people of Merseyside continue to support it. They will not do this if all the recent most regrettable and very negative publicity persists. (K4C)
- (345) We suggested to the ministry a long time ago that they should find a flock with salmonella and test the eggs for salmonella. They refuse to do this. There have been no scientific studies to determine whether the eggs themselves are infected. (A49)

In cases coded as ‘highly salient’ by the second coder, the antecedent-trigger is usually further away from the VPA, or especially long (346), or it appears in a less prominent syntactic position (347).

- (346) But the most controversial part of the bill relates to unofficial strikes, a legacy of the one-day summer stoppages on the London Underground. A union would have to issue written warnings to all those members it had learned were threatening an unofficial stoppage telling them that they risked dismissal without compensation if the strike continued. Mr Fowler conceded that a union might not always be able to do this but it would have to show that it had used its best endeavours. (AAL)
- (347) A home win is a treat yet to be enjoyed by manager Don Howe and is endangering the club’s proud record of 25 years in the First Division, a tenure bettered only by Arsenal, Everton and Liverpool. Most of those seasons have involved a struggle against relegation and, while they have been lucky to survive some, this team are good enough to stay up. They will do that if they win two of their remaining three fixtures at Highfield Road, but current form suggests that will not be easy. (AHU)

In two of the three occurrences coded by the second coder as ‘activated’, the antecedent is rendered less accessible by the mention of much more salient events or actions in the context. In (348) the club’s decision is only mentioned as part of the condition on the participation of a player; in (349) the antecedent-trigger and VPA are separated by an intervening sentence which changes the topic to the current economic situation, rather than the Chancellor’s policies. Both examples were also considered activated in the first coding. In such cases, as predicted by Gundel, it is clearly insufficient to retrieve the intended antecedent, requiring the use of a demonstrative pronoun.

- (348) If Steve doesn't play for Liverpool he is obviously unlikely to play for us. But if he and the club feel he is close enough to soundness to allow him to travel and let us decide, we will do that. (A8C)
- (349) If the Chancellor expects the private sector to lead the economy out of recession, then I'm afraid he's going to be disappointed. There's so much debt in private sector and it isn't enough to generate sustained growth. He could have used exports to get the economy going with a substantial reduction in the exchange rate and greater devaluation. We've had devaluation of 14 per cent and we would need another 10 per cent. He hasn't done this either. (CEK)

In the third occurrence, however, the antecedent is arguably in focus, since the proposition containing it is the antecedent of *so* in *I suppose so*, and it becomes the topic of the rest of the conversation. In this case, the activated coding was likely motivated by the distance between the antecedent-trigger and VPA, even though here the intervening turns reinforce the saliency of the antecedent rather than move it out of focus as in (349)

- (350) I happened to mention that a mutual friend was going to Bibury with his girl for the weekend. Teddy [Edward Heath] looked at me with horror. 'You don't mean to say they're sleeping together?' he whispered. 'I don't know. I suppose so,' I replied. 'Good heavens,' said Teddy, 'I can't imagine anyone in the Conservative Association doing that.' (AJY)

Thus, discrepancies between the two codings stem not so much from the choice of statuses—Gundel's cognitive statuses, or degrees of saliency as in the second coding—as from the way each coder understood the annotation scheme. The first coder closely followed the distinctions established by Gundel and analysed the examples in terms of their proposed categories, whereas the second coder tended to interpret saliency as a function

of the distance between the VPA and antecedent-trigger. It should also be noted that a coding of this nature is a difficult task, and requires competent coders provided with clear instructions if one is to obtain useful results.

6.6.4 Conclusion

The corpus study reported here brings mixed results regarding the role of saliency in VPA alternation. Gundel's model, in which prominence in memory and focus of attention are central factors in determining the saliency of a referent, correctly predicts that *do it* is mostly used with maximally salient antecedents, but it does not explain why *do this/that* are so commonly used with highly salient antecedents.

The results of the alternative coding, which relied on a gradient conception of saliency, seem to suggest that a highly salient antecedent is the norm for all VPAs, rather than a property distinguishing *do it* from *do this/that*. Unlike in the cognitive status coding, there is no evidence that *do this/that* are more likely to be used if the antecedent has low saliency. This contradictory finding may be partly explained by the fact that, as discussed above, the level of saliency in the second annotation of the data was assessed largely based on the proximity between the VPA and the antecedent-trigger as well as its syntactic prominence, with 'activated' being reserved for more distant or less syntactically prominent antecedent-triggers. Since this was rarely the case, the 'activated' status was used in just a few cases.

Chapter 7

Function of the VPA clause

7.1 Introduction

In previous chapters we have considered how VPA choice is determined by various features of the antecedent, such as agentivity, saliency, discourse-newness and the contrast between given/new information on a more general level. This chapter examines the role of the VPA-containing clause and more specifically the function it fulfils in referring back to the antecedent action, such as specifying further properties of the action by means of an adjunct (351), identifying the agent (352), or other.

(351) With the move to the new system, we're abolishing that subsidy, but we're doing it in two goes: 50% of it this year and all of it will go next year. (KRT)

(352) It was clear that funds were being embezzled but who could be doing it? Could it possibly be the accountant Edward Morris? (CKD)

As previously discussed, There is some empirical evidence that the structure of the VPA clause can affect VPA choice independently of the antecedent-trigger. Compare, for instance, the attested example in (353a)

with the variant in (353b).

- (353) a. He closes his eyes when he speaks and I don't trust anyone who does that (?this/#it). (AHF)
- b. He closes his eyes when he speaks. I hate when he does that/this/it.

In the original sentence, the VPA occurs in a relative clause modifying an indefinite pronoun, and *do that* is more natural than *do this* or *do it*. In the modified variant, the VPA appears in a time subordinate and has the same subject as the antecedent-trigger. Whilst *do that* remains acceptable, the decrease in acceptability for *do it/this* is not observed. These structural differences also mean that the VPA sentence fulfils a different function each time with respect to the antecedent: in the initial example, it denotes the general action of closing one's eyes while speaking, in the variant (353b) it refers to the same habitual action performed by the same subject.

In what follows we will consider how such variation in the way the VPA sentence refers back to the antecedent can influence pronoun choice. We will seek to identify the range of possible functions that VPAs may fulfil, and which functions are typically fulfilled by each VPA.

7.1.1 Background

The function of the anaphoric clause plays an important role in [Miller and Pullum \(2014\)](#)'s account of the choice between VPE and VPAs which was discussed in previous chapters. In the most typical 'auxiliary-choice' use of VPE, the context contains a salient choice of alternatives, and the VPE sentence is strictly limited to choosing one branch of the alternative. In the less frequent 'subject-choice' use, the context contains a salient property and the VPE utterance is limited to identifying who or what possesses that property. The relevant examples are repeated below:

- (354) He shops in women's.
 B: No, he doesn't. [COCA]
 [Compare #*No, he doesn't do it /this /that.*] (ex. 11, p. 9)
- (355) The boys cheered. I did too.
 [Compare #*I did it / this / that too.*] (ex. 20a, p. 13)

As discussed earlier, VPAs are preferred over VPE if the anaphoric utterance specifies additional properties of the antecedent action by means of a non-contrastive adjunct, such as in (356).

- (356) A: He shops in women's.
 B: He never does it alone. / He does it all the time. /
 He does it because that's the only place he can find things his size.
 #He never does alone. / #He does all the time. /
 #He does because that's the only place he can find things his size.
 (ex. 13, p. 10)

Miller (2011, 2013) also pointed to the role of the VPA clause in the use of VPAs specifically. He observed that the use of VPAs to introduce a further description of the action as in the *do it* variants of (356) is especially common with *do so*, which is followed by a non-contrastive adjunct in 83% of finite occurrences in Miller (2011)'s corpus. Another common use of *do so* observed by Miller is to temporally locate another event (357). According to Miller, this usage is encountered in all occurrences in his corpus where *do so* is not followed by an adjunct.

- (357) Nathan immediately bends down to pick them up but is jostled as he does so and stumbles, breaking his fall with his right hand...
 (COCA, Miller, 2013, ex. 4b, p. 124)

The above examples illustrate some possible functions for the anaphoric clause and how they play into the use of VPE, *do so* or *do it*. In what follows we will seek to identify further such functions and investigate their role in the choice between *do it* and *do this/that*.

7.1.2 VPA clause functions

The question underlying the coding is essentially this: what is the point of referring to the antecedent action using a VPA? By establishing the range of functions fulfilled by *do it/this/that*, we will attempt to identify, in turn, broad differences in usage that can further explain the alternation between them. A total of 11 such functions—including those defined above by [Miller and Pullum](#)—have been identified in the corpus (though ‘complement-other’ is in reality a mixed category); ‘other’ is used for occurrences whose classification remains uncertain at this stage. [Table 7.1](#) shows the frequency of the different functions, which are defined briefly in the list below. Further details of the classification method are provided in the relevant sections.

The full set of functions identified for VPAs is shown below in the order in which they will be discussed. Each function is defined briefly and illustrated with an attested example from the corpus.

The rest of this chapter will be organised as follows: first, we will examine a typical pattern where the VPA is followed by a non-contrastive adjunct, and VPE is dispreferred, as well as a related function where the VPA describes a different occurrence of the action with adjuncts such as *before* or *again*. We then review a series of functions more closely associated with VPE, namely signalling a new choice of polarity for the antecedent clause ([Miller and Pullum \(2014\)](#)’s ‘auxiliary-choice’), or identifying someone or something possessing a salient property ([Miller and Pullum](#)’s ‘subject-choice’). It will be shown that VPAs may fulfil similar functions to VPE, but do so less frequently and in very different discourse contexts. Separate treatment will be made of the relatively common use of VPAs with modals. We will then examine the case of VPAs functioning as complements to non-modal verbs, nouns or adjectives. After a brief survey of the use of VPAs in imperative clauses, we move on to a range of cases where the VPA occurs in an adjunct. We first discuss the use of VPAs in

purpose adjuncts to express an ‘intended action’, as we call it, and in time adjuncts to temporally locate other events or situations, before examining other types of adjuncts. The last function considered is the relatively infrequent use of VPAs as sentence subjects, rather than as predicates or complements. Lastly, we consider a small number of occurrences where the function of the VPA is unclear or not readily classifiable into one of the existing categories.

Adjunct-choice The VPA sentence describes further properties of the antecedent action, typically by means of a non-contrastive adjunct:

(358) [T]he aim is to encourage children, obviously, to remain nonsmokers and we do that by communicating directly with them erm, in their homes...(FLM)

Different occurrence of action The VPA sentence is limited to denoting another occurrence of the action token denoted by the antecedent-trigger.

(359) Godfrey always complained about her though they split up a year ago. He was doing it again. (AOR)

Polarity-choice The main function of the VPA sentence is to choose one branch of a salient polar alternative in the context.

(360) I spend a lot of time gloomily thinking I couldn’t possibly write about that, and then eventually I do it. (KRF)

Agent-choice The main function of the VPA sentence is to identify the agent of the antecedent action:

(361) ‘I’m hungry. I’ll make breakfast.’
‘Don’t worry, I’ll do it.’ (AE0)

CHAPTER 7. FUNCTION OF THE VPA CLAUSE

Modal-choice The main function of the VPA sentence is to express a modal judgement on the antecedent:

- (362) I stopped crying now, but I feel a bit shaky so I look down at my feet and try to figure things out. That's what I do when I get worried – I try to figure it out. But I can't do it – all I can think about is Marie's face. (A74)

Complement The VPA is the complement of a non-modal verb, noun or adjective and is not used in one of the above-mentioned functions.

- (363) She considered writing to Luke in London – she had even taken notepaper out – but realized that it would be directly confronting Moran. She could not bring herself to do it. (A6N)
- (364) The night before the wedding, Cameron Nielson Jr burned his manuscripts, one by one. Alexia stood by and watched him do it. (ALJ)

Directive The VPA sentence occurs in an imperative clause expressing a directive (order, request, etc.)

- (365) Frau Nordern turned her face away. 'Don't do that, Helga,' Herr Nordern said. 'Look at me, please. Thank you. (A7A)

Intended action The VPA occurs in a purpose adjunct and denotes an action that is intended; the corresponding main clause describes what is done or required for the action to occur:

- (366) 'So you see, [...] it's not easy to know what you want. Because to do that, you have to know who you are. Who you really are. (AC4)

Temporal location The VPA sentence is used to temporally locate an event or situation relative to the event in the VPA sentence:

- (367) Then she yawned suddenly, like a cat, revealing two rows of white, even teeth, before she covered her mouth. She lifted her head as she did this, and seemed to look straight at him. (ANY)

Adjunct The VPA occurs in an adjunct other than a time or purpose adjunct:

- (368) Nigel began selling review copies—his own and any others he could cadge—to give a little boost to his income. He had been too lordly to do this at one time, preferring to give them as Christmas presents. (AC3)

Subject The VPA is the subject of the sentence it occurs in:

- (369) So that's what I do – I bring all my plants and the flowers Mr Frost gave me a couple of days ago. It was nice of Mr Frost to do that. (A74)

Other The function of the VPA is either unclear or not classifiable into any of the above categories.

- (370) There is a rich (pause) erm, history of women in Scotland and we're only just beginning to discover that and publish that, and I think that we will establish traditions by doing that. (FLL)

7.1.3 Data: frequency of VPA functions

The various functions defined above appear in Table 7.1 in decreasing order of frequency (in the interest of clarity, the frequency of each VPA

CHAPTER 7. FUNCTION OF THE VPA CLAUSE

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Adjunct-choice	286 (57.2%)	95 (47.5%)	50 (25%)	431 (47.9%)
Complement	52 (10.4%)	33 (16.5%)	32 (16%)	117 (13%)
Modal-choice	55 (11%)	11 (5.5%)	35 (17.5%)	101 (11.2%)
Agent-choice	49 (9.8%)	18 (9%)	16 (8%)	83 (9.2%)
Polarity-choice	22 (4.4%)	9 (4.5%)	28 (14%)	59 (6.6%)
Temporal	6 (1.2%)	14 (7%)	12 (6%)	32 (3.6%)
Diff. occurrence	15 (3%)	5 (2.5%)	2 (1%)	22 (2.4%)
Intended action	2 (0.4%)	8 (4%)	5 (2.5%)	15 (1.7%)
Adjunct	5 (1%)	3 (1.5%)	6 (3%)	14 (1.6%)
Subject	3 (0.6%)	2 (1%)	7 (3.5%)	12 (1.3%)
Other	3 (0.6%)	2 (1%)	3 (1.5%)	8 (0.9%)
Directive	2 (0.4%)	0	4 (2%)	6 (0.7%)
TOTAL	500	200	200	900

Table 7.1: VPA clause functions

in the function considered will be reproduced in the corresponding section). The most frequent function by far is the one we call ‘adjunct-choice’, where the VPA sentence specifies additional properties of the antecedent action. VPAs also commonly occur as complements, chiefly with modal expressions, but, as might be expected given the work of [Miller \(2011\)](#) and [Miller and Pullum \(2014\)](#), they are less common in typical VPE functions, as sentence subject, or in adjuncts and directives.

The most notable discrepancies between VPA contexts are observed with the adjunct-choice function, which favours *do it* (and to a lesser extent *do this*) over *do that*, and the polarity-choice function, which is much more typical of *do that* than either *do it* or *do this*. Data for the other functions will be discussed in the relevant sections.

We now discuss the results of the function coding in greater detail. For each function, we explain the criteria for classification and then provide further analysis of the attested data.

7.2 Adjunct-choice

7.2.1 Coding scheme

We propose to call ‘adjunct-choice’ (following [Miller and Pullum](#)’s ‘auxiliary-choice’ and ‘subject-choice’) the use of a VPA to describe additional properties of the action by means of a non-contrastive adjunct. Representative examples are given below in (372)–(373).

- (371) Motorways are dangerous enough place at the best of times. To actually stop on the hard shoulder is very dangerous. But to do it in lane three, to get out and change a tyre is beyond belief. (K6D)
- (372) Would it be best to accept another cup of tea before trying to climb the stairs? But she must pour it herself. She must do this with very great care, not spill it or make any sort of noise. (AEA)
- (373) We realise that there are many problems caused by people who aren’t able to park er in a restricted area er and we need to ensure that people are able to do that satisfactorily and safely. (KRT)

All occurrences where the VPA is followed by a non-contrastive adjunct are classified as adjunct-choice. This excludes adjuncts specifying a different occurrence of the action, such as *again* or *before*, as well as reflexives used with the sense *too/as well*, unless they are followed by another adjunct as in (375).

- (374) As for going off at tangents, my dear, I do it myself, hormone balance notwithstanding. (AN8)
- (375) You can’t leave here because no other place would tolerate you or be able to support you. You know that too and that’s part of the reason why you befuddle yourself with drink. Don’t get me wrong, I’ve done it myself in the past and seen just how far it got me. Absolutely nowhere, that’s where it got me. (ASN)

Note that the adjunct may also precede rather than follow the VPA, as in (376), where *only for the money* is the focus of the *it*-cleft.

- (376) — [W]hat were generally the motivation behind the girls becoming prostitutes, I mean was it
 — I would say a lot of it is money.
 — Mhm.
 — Erm and I know one that I do know now said it's certainly, she said, it's only for the money I do it, nothing else. (FY8)

Although (377) below may appear to show a different context, since there is no adjunct after the VPA, it is in fact very similar to the above cases: the point of the sentence following *do this* is to explain why the subject teaches creative writing during his retirement. The PP *for extra income...* is a purpose adjunct that could occur directly after *do this*; the meaning is essentially as shown in the variant (378).

- (377) Nice chap and he teaches creative writing in Glasgow, used to be a teacher then a teacher trainer and then I think took early retirement and he does this but made the point that he simply uses it for extra income for pleasure and interest [...] (KGK)
- (378) He teaches creative writing in Glasgow. He does this for extra income for pleasure and interest.

Also included in this category are sentences where the VPA occurs in a direct (379) or indirect interrogative (380) headed by a *wh*- adjunct such as *why* or *how*, or as the complement of a noun denoting a property of the action, e.g., *a way of doing it* (381).

- (379) — Well you can guarantee they're gonna fight tooth and nail over this job creation scheme, to stop it.
 — Mhm.
 — Why do they do it, Douglas? (HM4)

- (380) I went to Grenoble and discovered they had a small project on the outskirts, which had been a piece of wasteland; they ploughed it all down, replanned it, had a complete new housing estate, hotel, the lot, in less than five years, where the City Council would be thinking about which bit of land to use, what to put there, how to do it, and ten years later they might think about producing a plan, and ten years after that (that's twenty years on) something would appear. (KRL)
- (381) [T]hey might be able to test their ideas about how matter behaves at very high density by using cosmology, and that's very important because we have no other way of doing it. Also, it's worth adding, it's a very, very cheap way of doing it. (KRH)

These differ from examples like (372) in that the VPA clause does not specify the manner, cause, etc., of the action: (379) asks about the cause, while (380) presents it as unknown or unspecified, or, in the most common case, identifies some other action or event as being the cause, manner or other, for instance, in (381), cosmology is considered the only way to test how matter behaves at very high density. (382), with the VPA in an identificational reversed pseudo-cleft structure, in which *why* is a fused relative, is an even clearer example of this: the reason why the subject shaved his head is already given by the purpose adjunct *so he looks better*; the VPA clause makes it clear that looking like Yul Brynner is also part of the reason.

- (382) But he's shaved his hair off so he looks better, he thinks he looks like Yul Brynner. Right? And that's why he's done it. (HVE)

7.2.2 Data

As seen earlier in Table 7.1, adjunct-choice is the most frequent function in the corpus. The predominance of this usage (almost half of occurrences)

is consistent with the findings of Miller (2011, 2013); Miller and Pullum (2014), who argue that this type of context leads to a preference for VPAs over VPE. Although the adjunct-choice category is quite broadly-defined (since it includes all occurrences that contain an adjunct either before or after the VPA), it is nevertheless surprising that it is so frequent compared to other broad categories such as polarity-choice.

ADJUNCT-CHOICE	
<i>Do it</i>	286 (57.2%)
<i>Do this</i>	95 (47.5%)
<i>Do that</i>	50 (25%)
TOTAL	431 (47.9%)

Table 7.2: Frequency of the adjunct-choice function by VPA

Do it and *do this* are much more frequent than *do that* in adjunct-choice contexts (see Table 7.2). This is largely explainable by the lower frequency of adjuncts after *do that* than *do it/this*. However, the data for *do that* show that it is also atypical in adjunct-choice contexts even when not followed by an adjunct, suggesting that *do that* is in fact rare with adjuncts in general, regardless of their position within the sentence. The data in Table 7.3 confirm this tendency: less than 10% of *do that* occurrences not followed by an adjunct are of the adjunct-choice type.

VPA	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Adj-choice	61 (23.8%)	17 (14.8%)	15 (9.3%)	93 (17.5%)
VPA –ADJ.	256	115	161	532

Table 7.3: Frequency of the adjunct-choice usage for occurrences without a subsequent adjunct (VPA –ADJ.), as a proportion of the total sample for each VPA

Table 7.4 shows the distribution of the two patterns we have distinguished, either with or without a subsequent adjunct. The VPA + adjunct

7.3. DIFFERENT OCCURRENCE OF ACTION

case is dominant, but it is still less frequent with *do that* than *do this/do it*.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
VPA + Adj.	243 (48.6%)	85 (42.5%)	40 (20%)	368 (40.9%)
VPA – Adj.	257 (51.4%)	115 (57.5%)	160 (80%)	532 (59.1%)
TOTAL	286	95	50	431

Table 7.4: Frequency of the adjunct-choice use in cases where the VPA is followed by an adjunct (+ Adj.) and those where it is not (-Adj)

7.3 Different occurrence of action

7.3.1 Coding scheme

In this section we consider the use VPAs with *again* or *before*, where the VPA sentence refers to another actual or putative occurrence of the antecedent action at a different point in time. In (383), *do that* refers to past cases of Eleanor’s hinting at an affair; in (384) *do it* refers to possible future instances of lending money to the addressee.

(383) Eleanor started hinting at an affair. She had done that before.
(AC3)

(384) I should never have extended credit to you then, and I won’t do it again. (ALJ)

7.3.2 Data

As shown in Table 7.5, this usage makes up a little over 2% of the sample. *Do it* is mostly used to express repetition of the antecedent action with *again* (10 of 15 cases), whereas all but one occurrence of *do this*, as well as the two occurrences of *do that* have *before* as the adjunct.

DIFFERENT-OCCURRENCE	
<i>Do it</i>	15 (3.0%)
<i>Do this</i>	5 (2.5%)
<i>Do that</i>	2 (1.0%)
TOTAL	22 (2.4%)

Table 7.5: Frequency of VPAs expressing a different occurrence of the antecedent action

A search in the BNC for the patterns *do it / this / that again* (with all forms of *do* considered) and *done it / this/ that before* followed by a punctuation mark shows that *do it* is the most frequent in both cases, but it appears to be much more strongly preferred with *again*.

	<i>Do it</i>	<i>Do this</i>	<i>Do this</i>	Total
VPA + <i>again</i>	471 (83.1%)	73 (12.9%)	23 (4.1%)	567
VPA + <i>before</i>	73 (54.9%)	39 (29.3%)	21 (15.8%)	133
TOTAL	544	112	44	700

Table 7.6: Frequency of VPAs followed by *again/before* and a punctuation mark in the BNC

7.4 Polarity-choice

7.4.1 Coding scheme

‘Polarity-choice’ describes contexts where the focus of the VPA sentence is on whether or not the action took place, hence on a choice of polarity. Below are some examples of polarity-choice VPA:

- (385) So if you were thinking of becoming a super model, my suggestion is that er you don’t do it alright because it’s er it’s not worth it.
(HUV)

- (386) I believe that they [national characteristics] certainly exist because history has given to each contemporary people a cast of mind, a set of habits and attitudes in their conduct of their own affairs, in the state of their minds, in their attitudes towards others, which make them individual and peculiar. I think history has done this, and by history I mean everything which has worked through history to produce that result—geography, climate, agriculture, economics.
- (387) Mr Jackson tells me to shove my dirty clothes in a placky bag, so I do that. (A74)

This function is superficially similar to Miller and Pullum’s ‘auxiliary-choice’ (‘aux-choice’, for short), the first of two types of usage for VPE, which they define as in (388). However, aside from the fact that it does not necessarily involve an auxiliary, the type of usage we discuss here is not subject to the same specific requirements as aux-choice. We therefore propose to call it ‘polarity choice’.

(388) **Type 1: Auxiliary-choice**

FORMAL CHARACTERISTICS: The subject of the antecedent is identical with the subject of the PAE construction and the auxiliary is (at least weakly) stressed, signaling a new choice of tense, aspect, modality, or (in the most overwhelmingly frequent case) polarity.

DISCOURSE REQUIREMENT: A choice between the members of a jointly exhaustive set of alternative situations must be highly salient in the discourse context, and the point of the utterance containing the PAE is strictly limited to selecting one member of that set.

(ex. 10a, p. 8)

Discourse-wise, the central difference between aux-choice and polarity-choice is that the latter does away with the requirement of a salient choice of alternatives in the context. What matters here is the polarity-selecting

function of the target clause, rather than the saliency of the antecedent and more specifically of a choice between alternatives. Thus, amongst the examples cited above, (385) and (389), but not (387), are auxiliary-choice. There is also no requirement that the antecedent and target clause should have identical subjects, so that the polarity-choice category also includes certain cases which Miller and Pullum refer to as instances of ‘dual focus’, or simply ‘mixed cases’¹, e.g., (389):

- (389) I promised Marie I wouldn’t go out with Fullblast again. Nicks things, and if I did that, the police would come and take me away.
(A74)

As already pointed out in earlier chapters, Miller and Pullum argue that VPAs are normally dispreferred in aux-choice contexts. However, they add that

do that (and to a lesser extent *do it* and *do this*) can be used to suggest that the point of the utterance goes beyond a simple choice between the members of the set of alternatives, without explicitly indicating what is at stake. (p. 10)

Taking (390) as an example, they explain that replying with *do that* instead of VPE

might be taken to suggest that the person referred to as “he” has some other activity contrasting with shopping in women’s that she intends to discuss. Or, she might be suggesting, with appropriate intonation, an implicit evaluative comment of the ‘*Can you believe it?!*’ type. (p. 10)

1. Miller and Hemforth (2014) suggest that in mixed cases, one of the choices is usually subordinate to the other. Accordingly, we classify cases of this type as polarity-choice when the choice of polarity is dominant.

- (390) A: He shops in women's
 B: No, he doesn't (do that).
 (adapted from 11, p. 9)

Therefore, in the analysis of polarity-choice cases, we will consider the influence of such discourse effects in the use of VPAs in contexts otherwise licensing VPE.

7.4.2 Data

Table 7.7 shows the frequency of VPAs in polarity-choice contexts. Though it is not a very common use of VPAs, it is immediately apparent that it is marginal with *do it* and *do this*, while *do that* is much more typical. To a certain extent, this lends support to the above claim by Miller and Pulum regarding the possibility for *do that* to occur in VPE-type contexts, a property which has already been observed in several attested examples discussed earlier. However, insofar as the usage we have defined is somewhat different to aux-choice, the data reported here would suggest a more general tendency for *do that* to be preferred over *do it* and *do this* in polarity-selecting contexts.

POLARITY-CHOICE	
<i>Do it</i>	22 (4.4%)
<i>Do this</i>	9 (4.5%)
<i>Do that</i>	28 (14%)
TOTAL	59 (6.6%)

Table 7.7: Frequency of VPAs in the polarity-choice usage

Occurrences of polarity-choice VPA may be divided into those where VPE is dispreferred for discourse (or possibly grammatical) reasons, and those where it is just as acceptable as the VPA. The corpus annotation in its current form does not capture the possibility of alternation with VPE, so

that it is not possible to provide detailed statistics on the frequency of these two situations. However, the sample contains enough examples to allow for interesting analysis of the parameters leading to the non-use of VPE. We will first review a range of factors which make VPE dispreferred independently of the polarity-choice context, and then focus on those cases where alternation with VPE is expected to be possible (regardless of whether it actually is).

7.4.2.1 Factors preventing VPE

The primary obstacle to VPE in polarity-choice contexts is the absence of a salient choice of alternatives in the discourse context: if the intended antecedent of *do it / this / that* is not salient enough, it cannot be accessed by VPE. As seen in Chapter 6, the saliency of the antecedent may be reduced by, amongst others, the structure of the antecedent clause, the distance and syntactic structure of the antecedent-trigger, or the lack of a specific trigger (the ‘inference’ cases discussed in Chapter 5)². (391) shows how distance reduces saliency: *thought about* raises the question of whether or not Marie will tell the truth, but the intervening sentence between the antecedent and VPA clause makes this proposition much less accessible. If this sentence is removed (as well as the preceding *that*-clause, which also adds to the distance from the VPA) the antecedent becomes highly salient and VPE is felicitous.

(391) Marie had thought about telling him the truth: that Bella had seen him and the police would soon be after him. She wanted to hear his voice falter, to see the fear and confusion in his eyes. She realized though, that if she did this/#did, he would never come

2. Exophoric uses are another obvious situation where VPE is dispreferred due to low saliency, since as Miller and Pullum show, it is very rare for the extralinguistic context to make a choice of alternatives salient. However, the corpus does not contain examples of exophoric polarity-choice VPA

out with her to the Lock. (ACB)

- (392) Marie had thought about telling him the truth. She realised, though, that if she did...

A consequence of the antecedent being insufficiently salient is that VPE is potentially ambiguous: the relevant antecedent may be salient in its own right, but it competes with other more salient referents that are also potential candidates for the interpretation of VPE. Frequently, the most likely interpretation of VPE in context is implausible or even nonsensical. For instance, *if she did* in (391) is likely to be understood as *if she saw the fear and confusion in his eyes*, whereas in (393), *I have* would mean *I have been pleased*, rather than *I have stopped it*.

- (393) — In your years as a deputy, have you stopped production because safety was at risk?
 — E- yes, I have. Machinery unfit to do I've stopped it for that. Managers haven't been pleased but I've done it. (HMG)

(394) is an example with a nominal antecedent-trigger already cited in Chapter 5, where *do that* refers not the salient property of being a songwriter, but instead to the activity of writing songs, with the implication that this is not the speaker's profession. First of all, if the speaker wanted to deny that he is in any way a songwriter, professional or otherwise, VPE would have to be *I'm not* rather than *I don't*. More importantly, the noun *songwriter* cannot be interpreted as a 'concealed question' as in Miller and Hemforth's experimental materials; it does not make salient the question of whether or not the addressee writes songs.

- (394) — So what's what's your occupation Martin, are you a songwriter?
 — Erm well not as such as a songwriter, I mean I do that yes but I'm my occupation I'm a self-employed window cleaner. (HMD)

The choice of *do that* over VPE in (395) hinges on the change of subject from the antecedent to the target clause. Whereas the first utterance brings

up the question of whether or not the particular government at issue will change its mind, the *do that* sentence asks whether this can be expected of governments in general.

- (395) ‘[...] Once all this country is solid against the Act, and all across the Lowlands, then the government must think again.’
 ‘Do governments do that? [...]’ (AON)

7.4.2.2 Alternation with VPE

We now examine the examples of polarity-choice VPA that would be equally felicitous if VPE had been used instead. Most such examples are of *do that*, which is attested in contexts normally typical of VPE, such as conditional clauses (396)–(397) or or answers to polar questions (398) :

- (396) You can not will yourself to fight against the truth, he wrote. You cannot will yourself to hold out against reality. If you do that, he wrote, then sooner or later the will will crack and the truth will emerge, reality will re-assert itself. [Compare: *If you do,...*] (A08)
- (397) She went down the concrete steps. The nettles in the tunnel entrance had been partially flattened. She barged through them: if you did that they didn’t sting you. (ACB) [Compare: *If you did,...*]
- (398) — So do you actually allow snakes to bite human beings to [collect venom samples] — No, I’m glad to say that we don’t do that. What erm happens is that we obtain the toxin from erm from the from a serpentarium. (KRF)

The use of *do that* here illustrates the effect described by [Miller and Pullum](#) where the VPA sentence goes beyond a choice of alternatives. This is best shown by (398), where the speaker interrupts the addressee’s question to reject the suggestion that venom is collected by having snakes bite

humans, and goes on to describe the actual collection method. By using *do that*, the speaker is also distancing himself from a practice which would be regarded as dangerous and unethical. (397), similarly to (395), involves a shift from a specific situation in the antecedent clause to a generic one in the VPA clause, a contrast which appears to favour *do that* regardless of whether VPE is also possible. As for (396), the contrast between VPE and *do that* is less obvious, but it may be that *do that* is used to stress that the event is unlikely to occur.

Whilst *do that* can often freely alternate with VPE in its most typical use, this is much rarer with *do it* and *do this*. In particular, only one occurrence of *do this* has an antecedent that is sufficiently salient for VPE, namely (399). Just like the *do that* examples above, the VPE sentence goes beyond simple polarity-choice and signals the speaker's intention to continue discussing the topic, which is exactly what Strauss (2002) described for determiner or pronominal *this*.

(399) I said, Well we're spending all this money on outings, we could buy a bungalow at the seaside and let them all go you know pensioners go in their turn free. And so we did this (pause) and we had this bungalow at Mablethorpe which is still running, we've been down, we went down a few weeks ago. And the pensioners go down to this bungalow free, and it's in lovely spot and we gave something like two thousand pounds for it. [Compare: *And so we did...*] (FYH)

Do it seems more likely in aux-choice contexts if the antecedent-trigger is a VPA (400). In some occurrences, it conveys a degree of surprise on the part of the speaker, as suggested by *actually* and *bloody* in (401). In this case, it may be argued that the purpose of the *do it* sentence is not so much to confirm that the subject bought jogging shoes, but to point out that there is something striking about his decision to invest in sporting apparel and equipment, for instance because he normally avoids exercise.

CHAPTER 7. FUNCTION OF THE VPA CLAUSE

This type of implicature is specific to *do it* and is also found in contexts that do not license VPE because the antecedent is insufficiently salient, such as (402).

(400) [News report: British coal miners were ordered to under-report workers' dust exposure levels in order to avoid sanctions]

Miners know the pressure is on to increase the British Coal's competitiveness and save their jobs. But the sanctions imposed on a pit with high dust readings could be serious. The mine district could be shut down for three months.

[Cut to interview of mine worker] Basically we were told to do it, and if we didn't do it we'd got us money dropped. We had threats of being moved or we were sacked. (HMG)

(401) So you bought some jogging shoes. You actually did it. You bought some bloody jogging shoes (not just jogging shoes, with air-suspension heels, but shorts, two T-shirts, two pairs of running socks, a sweatshirt, a book on dieting and exercise, and a rowing machine/fitness centre for wet days costing £117.85). (ASD)

(402) She knew it was insane and dangerous, what she was doing. She knew it so clearly that three weeks ago she had made an appointment to take her breast lump to a doctor. But she had done it in a very peculiar way: she had booked in to a private Well Woman Clinic under an assumed name. The appointment was now for tomorrow. She was amazed at herself, but she had done it, and now in the early hours of the morning she did not know if, even so, she would go. (A6J)

Nevertheless, in the following examples of *do it*, the VPA does not seem to have any other function than choosing an alternative. VPE is possible and perhaps preferable in (404):

(403) The woman in the shop's pretty friendly and she tells me what to

do with the plants – you know, water them once a week, or keep them out of draughts and that. The other week she tried to get me to buy some different plants, but I didn't do it. I get the same plants every week – these real nice ones with the soft petals. [Compare: but I didn't] (A74)

- (404) I spend a lot of time gloomily thinking I couldn't possibly write about that, and then eventually I do it. [Compare: eventually I do] (KRF)

7.5 Agent-choice

7.5.1 Coding scheme

In this section we will be looking at cases where the focus of the VPA sentence is on who does the action, rather than when / how / why it happens, or whether or not it happens. (405) is a very representative example: the speaker makes the point that smoking is common by generalising from people at the party to 'everybody', and then giving examples of various people who smoke. Similar examples are shown in (406)–(407).

- (405) 'There were people taking drugs at that party, I could smell it.'
 'That was patchouli oil, Mom, can't you tell the difference? God, you're so naïve, I can't believe it. Anyway, so what if there were people smoking a little—everybody does it, the teachers in school do it...'
 'I don't believe it, not at your school.' [...]
 'There are plenty of other schools in town. Ryan says a teacher tried to score off him. And the soldiers in Vietnam do it...'
 'I suppose Ryan does it?' Jo paused. If she admitted that Ryan smoked her parents would probably break them up and she didn't want that. (APU)

- (406) Well I think a lot of evaluation does go on. I mean I think part of being a proper professional means that one does attempt to evaluate what one is doing, and I think that most teachers do do this. (KRH)
- (407) It struck her that they [estate agents] made their money very easily. She had to pay for the lavish colour brochure (the details of which contained many mistakes), she provided the historical background and ran her feet off showing people over: all they did was make appointments (underpaid girls did that) and yet she had to pay them thousands. (ABW)

The function of the VPA clause here is similar to the subject-choice use of VPE, which [Miller and Pullum \(2014, p. 8\)](#) describe as follows:

FORMAL CHARACTERISTICS: The subject of the antecedent is distinct from the subject of the PAE construction, and stressed if it is a pronoun.

DISCOURSE REQUIREMENT: A particular property must be highly salient in the discourse context, and the point of the utterance containing the PAE must be strictly limited to identifying something or someone possessing that property.

Although [Miller and Pullum](#) observe that *do it / this / that* are dispreferred in typical subject-choice cases (408), the examples given in (405)–(407) appear to be valid cases of subject-choice VPA: all have distinct subjects in the antecedent and VPA clauses, and all conform to the discourse requirement specified above, with the antecedent clause introducing a salient property and the VPA sentence identifying someone possessing that property. Consider for instance (405): asserting *there were people smoking* makes the property of smoking salient, and the successive occurrences of *do it* identify various people who smoke, in other words, have the property of smoking.³

3. ‘Property’ is used here in the logical sense of function from entities to truth values.

- (408) a. The boys cheered. I did too [vs. #*I did it / this / that too*].
 b. She didn't say anything, and I didn't either. [vs. #*I didn't do it / this / that either*]
 (ex. 20a–b, p. 17)

Though it is not immediately clear what distinguishes these examples from the invented ones in (408), note that (405) does not contain *too*, which would be required if VPE were to replace *do it* in all occurrences other than the first (e.g., *the teachers in school do too*).

All examples considered in this section share the same identificational function as those just examined. However, many of them do not, in fact, meet the full set of requirements for subject-choice: (409) lacks contrastive subjects (since interrogative *who* refers to the unexpressed agent of *funds were being embezzled*), while (410) has contrastive subjects, but not a salient enough antecedent for VPE.

- (409) It was clear that funds were being embezzled but who could be doing it? Could it possibly be the accountant Edward Morris? (CKD)
- (410) Boys need to be discouraged from assuming that they know what women's position is. I think it's, it really begins in the home, this, because mothers can do quite a lot in not educating their own sons to think of them as servants. A lot of women do this, they're very tolerant about boys' mess in the home and untidiness generally, and in a sense they, they lay the foundations, right from the very beginning, of boys' growing up to think of women as kind of household servants. (KRH)

Note that in addition to cases like the above, we also include examples where the VPA occurs in a restrictive relative in which doing the action is the relevant property identifying the referent of the head:

This is also sometimes called an open proposition (cf. Prince 1986).

(411) A business relies on a driving personality taking it forward, and the confusion about getting somebody on board who is working with you in the business and who is putting money on board, is that they feel to some extent, proprietorial to that business, and you can find that that will result in those individuals who have done that tearing apart the business because they are trying to go in different directions. (HMH)

Since the usage considered here often deviates significantly from subject-choice as conceived of by Miller and Pullum, we will instead refer to it as ‘agent-choice’, a term which simply captures the function of the VPA clause without any implications as to the saliency of the antecedent or the non-coreference between subjects.

7.5.2 Data

The frequencies reported in Table 7.8 suggest that the agent-choice function does not favour any particular VPA.

AGENT-CHOICE	
<i>Do it</i>	49 (9.8%)
<i>Do this</i>	18 (9%)
<i>Do that</i>	16 (8%)
TOTAL	83 (9.2%)

Table 7.8: Frequency of VPAs in the agent-choice use

If we consider (405)–(407), we see that the choice between VPAs is largely explainable on the basis of independent properties, for the most part identified in earlier chapters. In (405), *do that* could be used instead of *do it* in the first occurrence (*Everybody does that*), but *do it* is preferred in subsequent occurrences since the antecedent is highly salient and becomes the main topic of the conversation. In (406), *do it* / *that* are acceptable

alongside *do this*, but the latter is motivated by reference to the current situation regarding teachers' evaluation of their job. The use of *do that* is in (407) may be explained by the implicit contrast in polarity (it is implied that estate agents did *not* make appointments as initially suggested) as well as the speaker's opinion that simply making appointments does not justify the fees charged by estate agents.

Since it has been argued that the examples just discussed are actual instances of subject-choice in the sense intended by Miller and Pullum, one would expect VPE to also be acceptable. The variants in (412)–(414) show that this is the case (albeit with some changes to the target sentence). (405) allows VPE if, as suggested above, it is followed by *too* in occurrences after the first. (406) has a non-finite VPA, but VPE becomes felicitous if emphatic *do* is removed. In (407), the repair makes VPE harder to accommodate, but it can be improved by simply adding a phrase that makes this correction explicit:

- (412) Anyway, so what if there were people smoking a little—everybody does, the teachers in school do too
- (413) I think part of being a proper professional means that one does attempt to evaluate what one is doing, and I think that most teachers do.
- (414) [A]ll they did was make appointments (or rather, underpaid girls did)...

The acceptability of VPE here lends further support to the claim made by Miller and Pullum that VPAs may sometimes be used in contexts otherwise typical of VPE to achieve certain specific discourse effects. This possibility appears to be limited, however, since VPE is often dispreferred due to the low saliency of the antecedent (415), or other independent factors, such as the use of the progressive in the VPA clause but not the antecedent clause (416).

- (415) Peter waited until they had left the spectators behind-before he risked speaking. ‘Do you want to *open the wine?*’ Kate nodded, and rummaged in the carrier bag for the bottle and the corkscrew. ‘Can you manage? I’ll do it/#I will if you want.’ (CKB)
- (416) It’s widely known—because the unions have publicized it—that the Department is looking at bailing out Huerter Textiles. Henry and I are doing it/#are, as the resident textile rescue-squad. You just haven’t heard about it because you’ve been too busy. (AB9)

Additionally, in agent-choice cases, VPE is infelicitous with syntactic mismatches (as already shown by [Kertz \(2008\)](#) for voice mismatches), in exophoric uses⁴, and when followed by reflexive adjuncts having the sense of *too* (418).

- (417) ‘I hear your husband was blackballed when he tried to join the Country Club. Wonder who did that/# did?’
Mrs Doran stiffened. ‘That’s not true!’ she said, but her voice betrayed her.
- (418) His son is standing near the main entrance, ready for the journey, carrying a stick, a spear and a small leather bottle, similar to the sort of thing drunken youths use to squirt wine down their throats before braving the bulls in Pamplona. (I have done this/#have myself although no bull came anywhere near me.) (FAJ)

7.5.2.1 Agent-choice with modality contrasts

A notable exception to the preference for VPE in true subject-choice contexts is where the VPA occurs after a modal auxiliary and there is none

4. Note that [Miller and Pullum](#)’s attested examples of exophoric VPE are all of the auxiliary-choice type. They attribute the impossibility of exophoric subject-choice VPE to the fact that the extralinguistic context cannot, in general, make a property salient enough for VPE. They nevertheless show that it is possible to imagine situations where this is the case, and VPE becomes felicitous. See p. 21–22 of their paper for discussion of invented examples.

in the antecedent-clause. As (419)–(421) illustrate, this gives rise to a preference for VPA over VPE.

- (419) ‘Sara, how about being useful? Making the salad?’
 ‘I’ll do it/#I will.’ Veronica rose from her chair. She washed a Cos lettuce, leaf by leaf, then patted each of them carefully with a tea towel. (A0N)
- (420) My own view is the President and the other nations are to help Saddam wind up his regime, I would hope the Iraqis would do that/#would, but if not, they, then we, that is the United Nations. (KRU)
- (421) ‘Pertwee told me to wait for Charlie Hatton,’ Cullam said.
 ‘I wouldn’t do that/#wouldn’t. God, I was sick of him and his money.’ (A73)

The decrease in acceptability for VPE here cannot be due to the antecedent being insufficiently salient, since the auxiliary-choice variants below are acceptable (the auxiliary-choice reading of (423) implies that Sara is replying, rather than Veronica, as indicated by the alternative continuation *Sara rose from her chair*).

- (422) My own view is the President and the other nations are to help Saddam wind up his regime, and I hope they will.
- (423) Sara, how about being useful? Making the salad?
 ‘I will’. Sara rose from her chair...
- (424) Pertwee told me to wait for Charlie Hatton, but I won’t⁵.

The corpus contains just two counterexamples where VPE is equally felicitous (425–426):

- (425) Don’t knock yourself. I mean, you use words like ‘incandescent’ and ‘eminently’. I wish I could do that/could. I’d really like to be

5. *I wouldn’t (do that)* is unnatural here if uttered by the same speaker

able to speak like you, you know. You may be homeless but at least you're not a dumb blonde like I am.' (AOF)

- (426) I think what what what never ceases to amaze me about about people in the flats on benefit, is is how many of them do manage, how many of them never owe a penny. How many of them turn their kids out, erm you know, in decent clothes, properly fed. I mean I don't think I could do it/could. I I find it extraordinary. (FY7)

The decisive factor licensing VPE here is that there is an implicit contrast in polarity between the antecedent and VPA clause (as opposed to an explicit one in e.g., (421)). The apparent constraint against subject-choice VPE after modals does not apply, however, if there is a contrast in polarity such that the antecedent is negative and the target clause is affirmative:

- (427) 'Civilisation,' he cried. 'Get him civilised. If you don't do it, I will. Using the proper knife and fork, nice haircut, doing what he's told, all that kind of thing.' (AMB)

Lastly, (428) illustrates an analogous problem to the one discussed here, with an aspectual rather than a modal auxiliary. This example satisfies the criteria for subject-choice (*It would have been a mad gamble* makes the choice between invading and not invading salient (aux-choice) and the VPA sentence identifies one country that launched an invasion of the sort described), yet VPE is dispreferred. This appears to be due to the choice of auxiliary: VPE becomes felicitous if the past perfect is also used in the antecedent clause, as shown in the simplified variant (429). These data suggest that VPE is also be dispreferred in subject-choice contexts if it introduces a contrast in aspect. Contrary to modality, however, there is insufficient evidence in the corpus to confirm this hypothesis at present.

- (428) Whichever foreign power captured Ireland held a pistol aimed at Britain's heart. It would be a mad gamble of course, to launch

an invasion from a European port and hope to evade the British fleet. But the French had done it/#had. In the autumn of 1798, three frigates carrying more than a thousand soldiers had sailed into Killala Bay. (B1X)

- (429) It would have been be a mad gamble to launch an invasion, but the French had.

7.5.2.2 Agent-choice and polarity contrasts

We now briefly consider agent-choice cases also involving a contrast in polarity, similar to (427) or (421) above. Analysis will be limited to those occurrences that satisfy the discourse requirements for subject-choice and auxiliary-choice jointly, which Miller and Pullum refer to as ‘dual focus’ or simply ‘mixed cases’ (fn. 8, p. 8). The properties of such mixed cases remain unclear—Miller and Pullum leave open the question of whether they are also subject to their proposed conditions on VPE.

Preliminary evidence suggests the acceptability of VPA in mixed cases varies, compare (430)–(431), both of which have VPE in the original sentence:

- (430) The boys cheered, but I didn’t #do it/this/that. (adapted from Miller 2011, ex. 3b)
- (431) Presidents don’t *write policy and spend money*, the Congress *does (it/this/that)*. (adapted from Miller 2011, ex. 7a)

While modality contrasts in subject-choice contexts are quite common with VPAs, polarity contrasts are much rarer: the corpus contains just two occurrences in addition to (427); they are shown below in (432)–(433).

- (432) ‘Ryan’s dad smokes, Mom,’ Jo answered at length. ‘Everybody does it.’
‘Your father and I don’t do it, Jocasta, and neither do you. (APU)

(433) ‘Oh, I say!’ he cried. ‘Don’t tell me that you are going to get married! Well, it mustn’t be till July! I am booked up solid.’ His soft mincing voice dwelt fractionally too long on every vowel.
 ‘No, we are not getting married,’ Catherine reproved him.
 ‘I am glad...Shouldn’t say things like that, I suppose,’ he simpered, ‘but all my friends are going off and doing it. There will not be anybody of our age around here soon.’ (ANL)

(432) is the continuation of an earlier example (405) and is a typical case of using *do it* for multiple references to a highly salient and topical antecedent. While this topic-maintaining function is not incompatible with VPE, it is much more closely associated with *do it*. In (433), VPE is impossible as it stands since the VPA is a coordinate non-finite VP, but it becomes felicitous if there is no coordination (*all of my friends are*), despite the distance from the antecedent-trigger. It is still more natural if the antecedent-trigger is near, as in the variant below:

(434) We’re not getting married, but all of my friends are (doing it).

As regards (427), VPE may be used throughout (*If you don’t, I will*), but a VPE > *do it* chain seems less natural (*?If you don’t, I’ll do it*). Aside from the fact that *Get him civilised* is essentially a clarification of *Civilisation* just before, the use of *do it* here is the idea of ‘success’ associated with doing the action: the sentence may be paraphrased as something like *If you don’t to get him civilised, I’ll manage to do it*.

7.6 Modal-choice

7.6.1 Coding scheme

We label as ‘modal-choice’ occurrences where the VPA occurs after a modal auxiliary (435), verb (436), adjective (437) or noun (438), and the VPA

sentence is limited to expressing a modal judgement on the antecedent.

- (435) If the Europeans fight like hell over new investment and where European headquarters should be, maybe we should consider running everything from here in Detroit, the way it used to be. We could do it, of course. (AC2)
- (436) I didn't like to tell you before but I've learned to be a dab hand at tapping the boots myself. Had to do it see, dad's been complaining about me being lazy. (CKD)
- (437) John then goes on to say well of course we're going to refer this to the Secretary of State. Well of course he is because he is obliged to do it. He's obliged to do it because it is a departure a very significant departure from his own development plan the plan that he piloted through to to tell people exactly what the policies of North Yorkshire County Council were to be. (HMP)
- (438) — Well we don't believe that any animals should be made to perform these kind of silly tricks purely for the purposes of human entertainment.
 — I don't think it's silly.
 — They simply don't have the right to do that. (HV3)

The above definition of modal-choice means that not all occurrences of VPAs after modals are considered as modal-choice. For instance, the following is an adjunct-choice case, since the VPA is followed by an adjunct falling under the scope of the modal: the sentence conveys not just the possibility of communicating something important, but also specifies when and where it is possible for the subject to do so.

- (439) To his disgust he found that the workers would no longer respond to his call for mass meetings during working hours. They now told him, in no uncertain terms, that if he had something important to

communicate, he could do it in the plant car park during the lunch break, or after working hours. (AC2)

7.6.2 Data

7.6.2.1 Frequency of modals with VPAs

Modals are the second-most frequent embedding structure for VPAs in the corpus (27% of occurrences), and the most frequent when the VPA is non-finite (34%). *Do it* and *Do that* are the forms most often occurring after a modal (33% and 25% of all tokens respectively). Table 7.9 shows the distribution of the modal-choice use across VPAs. It is most frequent with *do that*, less common with *do it* and quite rare with *do this*.

MODAL-CHOICE	
<i>Do it</i>	55 (11%)
<i>Do this</i>	11 (5.5%)
<i>Do that</i>	35 (17.5%)
Total	101 (11.2%)

Table 7.9: Frequency of VPAs in the modal-choice use
(% of total sample for each VPA)

Given that VPE is at least formally possible after modals, the question is then what motivates the choice of a VPA over ellipsis. (435) appears to satisfy the requirements for auxiliary-choice: *maybe we should consider running everything from here in Detroit* makes salient the question of whether or not this situation will arise, so that VPE is felicitous (*We could, of course.*). *Do it* may be favoured by the previous occurrence of VPE (*the way it used to be*) and the ‘success’ meaning (the interpretation being along the lines of *we could do it successfully*). VPE is also acceptable in (436) if the subject pronoun is not omitted and if there is a pause after the VPE (*I had to, see...*). In (437), previous anaphoric reference to the antecedent with VPE

(*of course he is*) then *do it (he is obliged to do it)* seems to strongly favour the choice of *do it* in further mentions of the action. As for (438), the antecedent is too distant and therefore insufficiently salient for VPE.

7.6.2.2 Modal forms and meanings

The corpus coding includes details of which particular modal the VPA occurs with, but does not distinguish between dynamic, epistemic, deontic modality and so on. We will therefore concentrate on the frequency of different modals within the modal-choice sample, and then make some further comments on the modal meanings encountered. First of all, it is necessary to provide some preliminary information on the frequency of modals generally. Biber et al. (1999) recorded the frequency of various modal verbs and expressions (e.g., *had better*), based on data from the LSWE corpus. For auxiliaries, they give the following order of frequency, from most to least frequent (see their figure 6.8, p 486): *will, would, can, could, may, should, must, might, shall*. They do not provide a similar ranking of modal verbs, but this can be inferred from the frequencies reported in their table 6.6. (p. 489), namely (in decreasing order of frequency): *have to, (had) better, (have) got to, need to*.

In the case at hand, we will be taking as a baseline the frequencies of modal auxiliaries and verbs in the entire BNC. This is because Biber et al.'s reported frequencies are approximate (expressed in terms of e.g., more than 200 / less than 100 tokens per million words) and also comprise both British and American English, while our study is limited to British English. The frequency of modals⁶ in the BNC are shown in Table 7.10 alongside those in the 900-occurrence corpus of VPAs, and finally in the modal-choice cases only. Some caveats: first, we disregard modal adjectives and

6. Note that we do not consider *going to* as a modal verb, as Biber et al. do. Moreover, *be to*, which is attested only once in our sample, is excluded from the count as it would be very difficult to restrict the BNC results to modal uses.

CHAPTER 7. FUNCTION OF THE VPA CLAUSE

nouns, which [Biber et al.](#) do not include in their table (except for *supposed to*), so that we have no information on which modal nouns or adjectives are the most frequent in general. Secondly, there is some noise in the data for modal verbs, as the [VP]*to* may also be a purpose adjunct (e.g., *[H]e can't get no money, cos he's used what money he had to pay off the card.* (KCT)). This means the frequencies are slightly overestimated (though it can safely be assumed that the irrelevant cases are far less frequent than the modal ones).

MODAL	BNC	VPAS	MODAL-CHOICE
<i>Would</i>	242195 (18%)	62 (25.1%)	16 (19.5%)
<i>Will</i>	239878 (17.8%)	33 (13.4%)	13 (15.9%)
<i>Can</i>	228485 (17%)	50 (20.2%)	21 (25.6%)
<i>Could</i>	158325 (11.8%)	52 (21.1%)	17 (20.7%)
<i>May</i>	111748 (8.3%)	0	0
<i>Should</i>	107822 (8%)	11 (4.5%)	5 (6.1%)
<i>Have to</i>	79842 (5.9%)	18 (7.3%)	3 (3.7%)
<i>Must</i>	69293 (5.1%)	9 (3.6%)	3 (3.7%)
<i>Might</i>	58344 (4.3%)	3 (1.2%)	1 (1.2%)
<i>Shall</i>	19407 (1.4%)	3 (1.2%)	1 (1.2%)
<i>Need to</i>	14844 (1.1%)	2 (0.8%)	1 (1.2%)
<i>Got to</i>	8921 (0.7%)	2 (0.8%)	1 (1.2%)
<i>(Had) better</i>	2932 (0.2%)	2 (0.8%)	0
<i>Ought to</i>	4970 (0.4%)	0	0
<i>Total</i>	1347006	247	82

Table 7.10: Frequency of modal auxiliaries and verbs (i) in the BNC, (ii) in the VPA corpus, (iii) in the modal-choice sample

The overall frequencies in the BNC (shown in the first column) are very similar to what [Biber et al.](#) described, with the possible exception of *(had) better* which appears to be somewhat less frequent than in the LSWE corpus. The frequencies observed with VPAs mostly conform to the general tendency, with the set *will/would/can/could* being the four most frequent forms, but there are some notable discrepancies. The past

forms *would/could* are proportionally more frequent than they are overall in the BNC, and also more frequent than their present-tense counterparts *will/can*. This is probably attributable to the typical use of the past tense in narration in the fiction sample. More striking, however, is the (near-) absence of *may/might*, which are otherwise common in the BNC (especially *may*, which is the fifth most frequent form). This is all the more surprising seeing as other forms that are less frequent overall are attested with VPAs in our sample (even if rarely), e.g., *must* or *shall*. The reason for this is not immediately clear, but it may again be an effect of register: Biber et al.'s data show that *may* is rare in fiction and conversation, but extremely common in academic writing. Even so, the separate corpus used in Chapter 4, which incorporates a wider range of sources including academic texts, contains just three occurrences of a VPA after *may*, such as (440). This would suggest that VPAs are not usually associated with the expression of epistemic modality. Finally, the distribution of forms in the modal-choice category largely reflects the tendency observed with VPAs in general, but *can* is more frequent than in the BNC or VPA data.

(440) Conflict: The bank could then make imprudent loans to Company X to keep it from failing. It may do this to prevent the securities affiliate from being sued by those who invested in Company X on the basis of its negligent advice. (ECD)

Detailed analysis of the modal meanings typically associated with different VPAs would require additional coding of the corpus. Discussion will therefore be limited to notable discrepancies between the modal-choice cases and the data from Biber et al.. Two interesting observations can be made: first, *can* almost exclusively expresses ability, despite the permission meaning (442) also being common in the LSWE data. Secondly, *must* is only encountered in the deontic sense, whereas the epistemic use dominates in the conversation section of the LSWE.

(441) What you've called the sloppiness of English is actually part of

its power. It means that, depending on the context, I can communicate something subtly different from what I intended before without us first having to go through the rigmarole of defining new terminology to extend the language. And this depends on our having very powerful and general rules in our minds for relating what is said to the broader context. Now, if we can do it, and it's not magic, then there must be some reasons we can do it, some rules we're following, and those rules can in principle be put into computers. (KRG)

- (442) The Race Relations Act says that you can do this you can have courses especially for black or Asian people [...] (HMA)

7.7 Other complements

7.7.1 Coding scheme

This section examines the use of VPAs in a range of non-finite complements, predominantly of the form *to*-infinitive, where there is no adjunct after the VPA, and the head is either a lexical non-modal verb (excluding e.g., *need to* or *have to*) as in (443), a non-modal noun other than *way* (which we considered as one instance of the adjunct-choice use), e.g., (444) or a non-modal adjective (445).

- (443) She grabs a hold of my arm and skips along beside me. She keeps poking me in the side with every step. 'Come on – skip!' I try doing it, but I get sort of tangled up. (A74)
- (444) Apart from the execution of the laws the executive has no authority to decide what shall be the relations between the federal government and South Carolina, any attempt to do this would be on his part a naked act of usurpation." (JSK)

- (445) After all, when you are ill, or when someone whom you love is ill, what you most want is someone to take the responsibility. Dr Dunstaple was very good at doing this. (EFW)

Occurrences in this category are similar to those classified under the modal-choice use (see Section 7.6), where the VPA is also a non-finite complement. The crucial difference is that the examples discussed here have little in common beyond the fact that they are complements, and the exact function of the VPA sentence as conceived of here will depend on the semantics of the head verb, noun or adjective. While occurrences may to some extent be classified into broader categories (such as verbs of wanting, requesting or other) there is no single semantic feature that would encompass all such categories.

Discourse-wise, the focus of the VPA sentence is whatever the head denotes with respect to the antecedent action, for instance *I try to do it* simply describes an attempt at performing the action, while *[He] was very good at doing this* refers to the subject's ease in doing the action. This is comparable to the modal-choice cases, but here there is a much wider array of possible meanings for the head than just rather than different flavours of modality.

There are three types of non-finite complements: *to*-infinitives, which allow VPE, and gerunds and bare infinitives, which do not.⁷ In practice, the availability of VPE in *to*-infinitival complements varies greatly depending on the head. For instance some verbs, such as *want* are well-known licensors of VPE, while with others VPE is possible but rare, e.g., *agree (I told him to stop but he wouldn't agree to)*. We will not go into detail about the extent to which different verbs, nouns or adjectives allow VPE after *to*.

7. A further possibility with some the verbs listed here is Null Complement Anaphora (NCA), where the entire *to*-complement is left out: compare *I tried to (do it) / I tried ∅*. The alternation between VPAs and NCA is left open for further study; see [Hankamer and Sag \(1976\)](#) and [Shopen \(1972\)](#) for discussion of the properties of NCA.

7.7.2 Data

This category is the second-most frequent in the sample, making up about 13% of occurrences. It is somewhat more frequent with *do this/that* (16% of tokens) than *do it* (10%). As shown in Table 7.11, the VPA is in most cases the complement of a verb; adjective and noun complements are less often attested. It is likely that this distribution of the different types of complements reflects the general tendency for English, not just the specific situation with VPAs.

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
Verb comp.	42 (80.8%)	26 (78.8%)	21 (65.6%)	89 (76.1%)
Noun comp.	5 (9.6%)	5 (15.2%)	6 (18.8%)	16 (13.7%)
Adj. comp.	5 (9.6%)	2 (6.1%)	5 (15.6%)	12 (10.3%)
TOTAL	52	33	32	117

Table 7.11: Frequency of VPAs in non-finite complements with no subsequent adjunct (excluding auxiliaries, modal expressions and *way*)

Detailed analysis of the semantics of the head for each complement type is beyond the scope of this study. Since very few of them occur more than once (the verb complement category, for instance, comprises as many as sixty-three different verbs), it is doubtful that VPA alternation is influenced by the specific meaning of the verb, noun or adjective under which it is embedded. The rest of the discussion will be limited to some notable tendencies emerging from the sample.

With verb complements, the most frequent verb aside from *go* in the idiom *going to* is *want* (447)–(448). The examples below further suggest that VPA choice has little to do with the meaning of the head verb; instead, we observe recurring features described in previous chapters, such as repetition of *do it* in (446), continuation of a current discourse topic with *do this* (447), and *do that* referring to actions performed by others but which the speaker would personally not engage in (448).

- (446) Er every night they used to play cards and when I was eighteen my father made me learn cards. He made me play cards, I'd played ordinary auction bridge at that time. But he made me do it, I didn't want to do it, but he he he made me do it. So we used to play auction bridge every night. (FXW)
- (447) 'Papa,' she said. 'We want to start a school.'
 'Really, my dear? Where?'
 'Here.'
 'But Charlotte, dear, we have no room. This house is full already.'
 'Oh, but we could change the house, papa. We could build a schoolroom.' 'Well, yes, I suppose so,' I said. 'But—why do you want to do this? Isn't it better to work as governesses, in some big fine house?'
 'Oh no, papa!' All three girls spoke at once.
 (FNY)
- (448) [W]hen I was a lot younger I had lots of er friends who er were all interested in the theatre, and some of them went into it professionally and some er some stuck with the amateur world as I did. But the ones who went into it professionally, I see them often on the telly advertising things and I don't really want to do that. (KRT)

VPAs functioning as noun complements are most often attested after *way*, but none of the nouns in the usage considered here occur more than twice. Commonly, however, the head noun denotes something which is either necessary or sufficient to carry out the action, e.g., *the money to do it* (449) can be read as 'enough money to do it'.

- (449) — But nevertheless, there's going to be a month period, plus there'll be five weeks where people won't be able to be treated for routine operations here.
 — Exactly so, because we haven't got the money to do it. (KRT)

CHAPTER 7. FUNCTION OF THE VPA CLAUSE

- (450) Well the answer is that residents' parking was offered in North Oxford I think about ten years ago, when it got to the top of the City's priority list. There was a public meeting, and the residents said we don't want it. The public meeting was held in August when the majority of people were away, I'm afraid, Bill. That was one of those things. They portray the fact that, yes, they come into public consultation—yes they do, not always at the right time. But things have changed since then. I said, there's loads of building going on and we have had money to do this and it's jolly well time. Actually, residents' parking is coming under review. (KRL)
- (451) Both parties, the man and the woman should always be in control of the situation and if for some reason they want things to stop, should have the authority and the power to do that. (K6E)

Lastly, adjectives attested with VPA complements mostly have to do with the subject's willingness or intention (or lack thereof) to perform the action, at least with *do it* and *do that* (452)–(453)

- (452) [M]y friend he used to come round here with his drugs, he used to tell me where they was you know. Probably thought that I'd go and get some but I, no way. That's one thing I want to try and keep off. You know being round here you might be tempted to do it but that's why I just keep in me flat and out of the way. (FY6)
- (453) Well I'm not prepared to become involved in er er discussing the speeches which I didn't hear or er which may well be out of context or anything of that sort. I'm not prepared to do that. (K6A)

7.8 Directive

7.8.1 Coding scheme

In this usage, the VPA occurs in an adjunct-less imperative clause and is not embedded under another structure. The point of the VPA sentence is to ask or invite the addressee to either perform or not perform the action (454)–(455). Although VPE is grammatically possible in imperatives, it is relatively constrained, see [Miller and Pullum \(2014\)](#).

(454) ‘[...] Why not switch your agents to checking those places?’
‘Do it/#Do,’ said Morgan.’ (ARK)

(455) Frau Nordern turned her face away. ‘Don’t do that/#Don’t, Helga,’
Herr Nordern said. ‘Look at me, please. Thank you.’ (A7A)

One example in the corpus, shown in (456), is formally identical to examples in this category, but was considered a case of subject-choice: the VPA sentence here does not just to express an order, but also emphasises that Philip must shoot the lamb instead of Lee (the subject pronoun *you* is likely to be carry contrastive prosodic stress, as indicated by the small capitals).

(456) Philip looked at where Rebel [a sheep dog] had separated one of the lambs from the flock. It hadn’t a chance. Lee raised the gun, aiming it. He fired, missing by miles.

‘I can’t,’ he said, turning to Philip. ‘I can’t see. My glasses are all steamed up.’

He was shaking from the shock of the recoil of the gun, which had sent him flying backwards.

‘YOU do it.’

Philip shook his head. He couldn’t do it.

(ABX)

7.8.2 Data

The corpus contains just 12 occurrences of VPAs in imperatives, most of them (7) with *do that*; *do this* is attested only once. 8 of these 12 cases match the particular usage considered here, with two occurrences of *do it* and six of *do that*. The *do that* cases provide clear examples of the speaker-exclusion meaning described in Chapter 6: all refer to an action that the addressee is performing or intends to perform (457); moreover, the negative form *Don't do that* can add a sense of reproach or warning to the directive (cf *urgently* in (457)).

- (457) Teversham nodded in acknowledgement. 'Miss Williams, we'll find the road again if we go straight and turn left by the hedge, won't we?' 'Don't do that,' McLeish said urgently. (AB9)

7.9 Intended action

7.9.1 Coding scheme

This category is exclusively comprised of VPAs occurring in non-finite purpose adjuncts headed by *to* (458) or, in one instance, *for* (459). Semantically, the VPA in the subordinate denotes a hypothetical action, while the main clause describes what is or should be done to carry out that action. The intended nature of the action is evidenced by the fact that in all cases the *to*-adjunct can be felicitously rephrased as *in order to* (e.g., *In order to do this, she unwrapped her legs...*). The *in order to* pattern itself is attested once with *do that* (460). Common to all such example is the impossibility of VPE after *to* in the adjunct (it is of course also impossible with the gerund in 459), in contrast to cases where the VP[*to*] is a complement.

- (458) Amanda inspected her drink. To do this she unwrapped her legs from each other, bent her right knee briefly and peered into the

empty glass. (AOR)

- (459) Why did he always have to work alone? What was so private about writing letters for business people, I who ought to have clerks for doing it? (CKE)
- (460) I think that the computer presents exactly that challenge and amongst the sorts of things I'm thinking of is that erm it's one thing to play with a computer toy, a game of some sort—we've all seen them in the bar and elsewhere—it's another thing entirely to devise your own game, to program your own rules in and then to bring your friend along and have them challenge it. Now in order to do that, you've really got to understand what a game is erm how you organise, for example, looking at the board, if it is a board game; how you're going to represent that in a program. (KRF)

7.9.2 Data

INTENDED ACTION	
<i>Do it</i>	2 (0.4%)
<i>Do this</i>	8 (4%)
<i>Do that</i>	5 (2.5%)
TOTAL	15 (1.7%)

Table 7.12: Frequency of VPAs in the 'intended action' use

As seen in Table 7.12, this usage is relatively infrequent, but somewhat more common with *do this* and *do that*. Note that *do it* is usually infelicitous in sentence-initial purpose adjuncts if it is not followed by an adjunct. As shown by (458) and the variant given below, this context is rather more typical of *do this* and *do that*, or even *do so*. No attested examples with *do it* are found in the corpus.

(461) Amanda inspected her drink. To do that / #do it / do so she unwrapped her legs from each other...

The Table 7.12 data suggest at least a weak preference for *do this* in purpose adjuncts with (*in order*) *to*. Moreover, it was already suggested in earlier chapters that sentence-initial adjuncts of the type in (458) especially favour *do this*. Although this preference is difficult to assess based on the small differences in frequency reported above, the relevant patterns can easily be retrieved by adding the appropriate punctuation to signal sentence boundaries⁸, namely:

```
. to do it|this|that| ,
```

which returns sentence-initial *to*-adjuncts followed by a comma;

```
in order to do it|this|that _y*
```

which returns *in order to* followed by any punctuation mark.

The data in Table 7.13 confirm both the prevalence of *do this* and the rarity of *do it* in the two types of purpose adjuncts examined. Further, they suggest that the preference for *do this* over *do that* is greater than the initial data might suggest, and that it is in fact the prototypical construction in the usage described here.

A look at the distribution of the two patterns across BNC sections reveals that they almost never occur in fiction and spoken discourse, and are instead encountered in magazines, various non-academic texts, and to a lesser extent in academic writing. Specifically, there is a tendency

8. The obvious limitation of these queries is that they ignore occurrences where the sentence boundaries are not marked by punctuation. However, since sentence boundaries not already encoded in the BNC, removing punctuation would also return occurrences with an adjunct after the VPA and, in the case of *to*-adjuncts, cases where the VPA is a subject (e.g., *To do this requires indigenous skills in chemistry* (B71)) or a complement (e.g., *a way to do this*).

7.10. TEMPORAL LOCATION

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	TOTAL
<i>To VPA</i>	1 (0.9%)	104 (89.7%)	11 (9.5%)	116
<i>in order to VPA</i>	1 (2.4%)	31 (75.6%)	9 (22%)	41
Total	2 (1.3%)	135 (86%)	20 (12.7%)	157

Table 7.13: Frequency of VPAs after *to* and *in order to* at sentence boundaries in the BNC

in explanatory or instructional discourse to introduce some putative action, and then start the next sentence by using *do this* in a purpose adjunct to describe how this action is carried out. Below are examples found in academic and non-academic texts from the BNC:

- (462) We need to be able to define a measure of strain which is valid for any type of deformation, not just extensional or shear distortion. To do this we consider the change in length of a small element dS of a line in the body as it is distorted to a new shape. (H0U)
- (463) The answer is, of course, to reduce the number of private motor vehicles. But to do this there must be viable, efficient and economic alternatives, either pleasant and safe facilities in which to walk or cycle, or reliable modes of public transport. (BN4)

7.10 Temporal location

7.10.1 Coding scheme

VPAs may be used in time subordinates to indicate that some event or state occurs before (464), after (465) or simultaneously to the event denoted by the VPA clause (466).

- (464) So we take our example we put it through these decoders. Before we do this we have to apply a mapping function M . (JP6, previous event)

- (465) [W]hen I saw my father glance disapprovingly at the lamp [...] I quickly lowered the wick. Having done this, I noticed all the more the effect of the pale light coming into the room. (AR3, subsequent event)
- (466) When they finish, the band pack up and fold their chairs up. There's lots of talking and laughing when they do that. (A74, simultaneous event)

Finite occurrences in this category, especially those in *when*-clauses, bear much resemblance to the polarity-choice cases discussed in (7.4): the VPA sentence may also be described as selecting a polar alternative, but this choice always serves as the basis for determining when some other event takes place. Note that these occurrences readily allow VPE, e.g., *before we do* (464) or *when they do* (466). With non-finite cases like (465), on the other hand, VPE is impossible, and the similarity with polarity-choice is much less obvious.

7.10.2 Data

TEMPORAL LOCATION	
<i>Do it</i>	6 (1.2%)
<i>Do this</i>	14 (7.0%)
<i>Do that</i>	12 (6.0%)
TOTAL	32 (3.6%)

Table 7.14: Frequency of VPAs in the temporal-location use

Miller (2013) observed that temporal location is the most frequent use of *do so* when it is not followed by a non-contrastive adjunct (though no figure is provided). It is comparatively rare with *do this /that* and a relatively marginal use of *do it*, regardless of whether or not an adjunct is

present after the VPA. It accounts for just under 4% of occurrences overall, as shown in Table 7.14, and about 7% of those that have no subsequent adjunct.

The corpus annotation does not currently capture the temporal relation between the located event and the event denoted by the VPA clause. cursory analysis of the data suggests that *do this* is predominantly non-finite and frequently occurs in the progressive with *while* (467), whereas *do that* is more commonly attested in finite clauses (the six *do it* occurrences are equally divided between finite and non-finite). Moreover, time subordinators containing *do that* are often analysable as causal (468), or in some cases, conditional (469).

- (467) She fried a rasher of bacon, a sausage and a slice of bread, then cut them up small enough for Eb to eat with a fork. Generally, he came downstairs to wash and shave while she was doing this, but today she took him up a tray so he could have his breakfast in bed. (FPM)
- (468) Sometimes she holds onto my arm in the street—you know, just natural, like I’m a girl and she’s just a friend or summat. I get real proud when she does that. (A74)
- (469) — So what you’re really saying is that the computers of the kind you’re describing are for the ordinary person. They’re for the child in school; they’re for the housewife; they’re for the businessman of the future, and an ordinary competent businessman has the capability and the possibility of actually learning how these devices work and using them, rather than just leaving it to the boffin, the scientist, the computer expert.
— That’s right, and until he does that, until he does, as it were, grasp the mettle and begin to express his purposes, his procedures, directly to the computer, he will always be at the risk that what gets expressed is not what he quite wanted, just a little bit differ-

ent, he and he alone, is the person who knows what the company purposes are. (KRF)

7.11 Adjuncts

7.11.1 Coding scheme

This category includes gerundial VPAs occurring in PP adjuncts as the complement of a preposition (usually *by* or *in*) (470) or as bare participial adjuncts (471). Like the complements discussed in the previous section, the adjuncts examined here have more in common syntactically than semantically: those in PPs are usually adjuncts of means or cause, but the function of participial adjuncts is more variable and not always easy to define.

- (470) I'd like to say I was a surrogate mother and I wasn't paid a penny for doing it! But I did it through the love, fo–, that I had for the couple, and that because they had waited sixteen year before they eventually found out they couldn't have children! (FLG)
- (471) No, Maidstone had been right all along: Sandison knew nothing about what went on in the city. He had naively stumbled into the middle of a very complicated and dangerous situation. Sandison desperately wanted to meet Elisa Stasi but he was not prepared to risk his life doing it. (ASN)

7.11.2 Data

The frequency of VPAs in participial adjuncts is shown in (7.15) Both types of adjuncts described above are attested about as often, with PP adjuncts being the more frequent.

VP[ING] ADJUNCTS	
<i>Do it</i>	5 (1%)
<i>Do this</i>	3 (1.5%)
<i>Do that</i>	6 (3%)
TOTAL	14 (1.6%)

Table 7.15: Frequency of VPAs in VP[ing] adjuncts

As regards PP adjuncts, by VP[ing] means adjuncts are found only with *do this / that* (472)–(473); *do it* is a less felicitous alternative in both cases⁹, and although it is infrequently attested elsewhere in the BNC in the pattern shown below, all examples are in fact cases of NP anaphora, e.g., [*Y*ou only improve at something by doing it].¹⁰ An example of *by doing it* from the COCA corpus is shown in (474).

- (472) This is an absolute nuthouse. If I said I saw sex organs when I looked at your rotten ink blots that would really get you going, wouldn't it? I'm not going to please you by doing that. You can keep them and your art therapy. (AC3)
- (473) I do not possess any pornographic publications or pictures of couples in lubricious postures, though I am aware that they exist to an ever more proliferating degree, and I have seen these things in

9. It is of course an open question to what extent the adjunct is actually responsible for *do it* being dispreferred here, as both examples combines a number of features independently favouring *do this/that*: in (472) the speaker is referring to an action just mentioned by the addressee and expressing his refusal to do as requested, and the antecedent event is referred to by *that* before the VPA. The facts of (473) are less clear, but *do this* refers to an ongoing smear campaign against the speaker and the following sentence considers possible reasons for spreading the rumours.

10. A search in the BNC for the pattern *by doing it|this|that|so _y** (where *_y** denotes a punctuation mark) returns the following results: 9 occurrences of *do it*, 19 of *do this*, 10 of *do that*, 49 of *do so*. The same query in the COCA returns 38 occurrences of *do it*, 130 of *do this*, 168 of *do that*, 387 of *do so*. This suggests that *do it* is possible in such adjuncts but less typical than *do this/that*. Moreover, many of the *do it* cases are NP anaphora, in which case only *do it* is possible

Robert's room and in my agent's office. The two people, in fact, who lead the campaign against me! Obviously, they are by doing this / #it attempting to discharge their own guilts, hide their own feverish obscenities, evade their own personal responsibilities and cleverly avoid gossip or rumour which would be more lethally accurate than that which they have generated themselves. (ADA)

- (474) In other words, the government should disseminate its information only when vendors cannot turn a profit by doing it. (COCA, *Washington Monthly*)

In-adjuncts as well as some of the participial adjuncts express simultaneity and (in most cases) causation; in that they are similar in that to the temporal-location use discussed in Section 7.10: *they feel good doing it* can thus be paraphrased as 'when / because they do it', for instance.

- (475) [TV debate: should bald men wear wigs?]
 Barbara at the end of the day [...] everyone's been saying if th– if they feel good doing it [= wearing a wig], and it gives them that much confidence and they feel great about it, whatever it looks like, why object? (HVE)

In contrast to PP adjuncts, participial adjuncts may also express simultaneity only, without any implications of causality (476). Despite their superficial resemblance with the temporal-location use, a *when* or *as* paraphrase is not always possible: (476) is to be understood not as *time when I did that* but rather as *time during which I was doing that*.

- (476) In any case, once I got more and more into the swing of writing, I started to spend more time at home doing that. (AOF)

Also, even if there is an obvious degree of overlap between *in*-adjuncts and bare participials, the former allow *do so* much more readily than the latter, as shown in this variant of (476): *#I started to spend more time at*

home doing so. On the other hand, alternation between *do it/this/that* in bare participials is relatively free, thus *do this* in (477) could felicitously be replaced with *do that*, for instance to suggest there is something unusual about the action, while *do this* seems to present it as more of a ‘novelty’ or perhaps ‘curiosity’ one might want to know more about.

(477) [Boy is watching television]

And then Boy cut back to the man on the bed [...] the man was saying I like your shoes, please take off your shoes ; and Boy cut backwards and forwards between this man and the politician beginning to lose his self-control and saying I would just ask people to forgive me really and to forgive my wife as well. Boy watched all of this as if he was seeing fragments of one, continuous and baffling programme. He would sit there all day doing this, sleeping sometimes in his chair but never tuning off, trying to make all these pieces of television fit together in some way. (AR2)

The difference in acceptability for *do so* depending on the type of adjunct it occurs in can be linked to Miller (2013)’s observation that ‘temporal location’ is a very typical use of *do so*: since this is the usual meaning of *in VP[ing]* cases, *do so* is allowed. Bare participial adjuncts, on the other express meanings more loosely related (or not at all) to temporal location, leaving a choice between *do it / this / that*, but normally not *do so*¹¹.

11. Three examples are found in the corpus used by Houser (2010) *do so*; they are reproduced below:

- (1) I’m assuming (and feel pretty comfortable doing so) that this put a crimp in their plans to eat them. [PXAngel03-8] (34j, p. 50)
- (2) We will only broach the technological adjacent possible at the rate at which we can make a living doing so. (p. 92)
- (3) [He] must be the only politician left in the House who avoids publicity and whose style is to follow the dictates of his conscience without making a spectacle of himself doing so. (p. 113)

7.12 VPA is sentence subject

7.12.1 Coding scheme

In this section we consider cases where the VPA is the subject the sentence it occurs in—either a VP[to] (478), a gerund (479), or a finite VP (480). In the case of (480), contrary to the first two, the VPA is not actually the subject, but is included in the subject (the subject of the embedded clause, *he*, is also part of the subject of the main clause). The impossibility of VPE after *to* in subject VPs is parallel to the situation observed with *to*-adjuncts of the type discussed in Section 7.9.

- (478) So that's what I do – I bring all my plants and the flowers Mr Frost gave me a couple of days ago. It was nice of Mr Frost to do that. (A74)
- (479) 'So what do we do? Phone up the London police?'
'No use doing that.' Keith shook his head. (AC4)
- (480) These white seagulls flash past, like flashing bits of light, and the traffic in the background rumbling away. It's like being at the sea, early in the morning with the seagulls and the sea pounding away and the fresh air. The old bloke's still chucking bread down. It's nice that he does that – you know, comes and feeds the birds like that. (A74)

7.12.2 Data

Do it/this/that rarely occur as subjects: the corpus contains 23 occurrences in total (2.5% of the sample), of which 13 in the pattern considered here where the VPA is not followed by an adjunct. The data in Table 7.16 might suggest a weak preference for *do that* if the VPA is the subject of the sentence, but this is likely due to the absence of an adjunct after the VPA;

7.12. VPA IS SENTENCE SUBJECT

SENTENCE SUBJECT	
<i>Do it</i>	3 (0.6%)
<i>Do this</i>	2 (1%)
<i>Do that</i>	7 (3.5%)
TOTAL	12 (1.3%)

Table 7.16: Frequency of VPAs as sentence subjects

the frequencies of VPAs in the entire sample are in fact very similar.

All examples have an extraposed subject as in (478)–(480), most commonly a VP[to]. VPAs in canonical subject position are only attested with adjuncts, see (481)–(482) (note that in both cases the antecedent-trigger is also the subject, and is syntactically parallel to the VPA).

- (481) Motorways are dangerous enough place at the best of times. To actually stop on the hard shoulder is very dangerous. But to do it in lane three, to get out and change a tyre is beyond belief. (K6D)
- (482) Escaping is bad enough but doing it in the middle of the night is inexcusable. (AMB)

The rarity of *do it / this / that* as subjects contrasts with the frequency of *do so*, which is virtually the default in this function. Table 7.17 shows the distribution of VPAs in three of the various possible configurations for extraposed or canonical subjects, described below with the corresponding BNC queries:

```
_y* to do it|this|that|so _v*
```

VP[to] subject in canonical position preceded by a punctuation sign and followed by a verb, e.g. *To do this requires indigenous skills in chemistry.* (B71)

CHAPTER 7. FUNCTION OF THE VPA CLAUSE

it [be] _j* to do it|this|that|so

Extraposited VP[to] subject with the pattern *It BE ADJ to VPA*, e.g., *It is impossible to do that*.

_y* _v?g* it|this|that|so _v*

Canonical VP[ing] subject¹² (same context as 1), e.g., *Lonie shook her head. Doing this dislodged the words stuck in her throat.* (GUK)

	<i>Do it</i>	<i>Do this</i>	<i>Do that</i>	<i>Do so</i>	TOTAL
VP[to]	0	11 (14.1%)	3 (3.8%)	64 (82.1%)	78
VP[to] extr.	15 (11.2%)	22 (16.4%)	18 (13.4%)	79 (59%)	134
VP[ing]	0	5 (38.5%)	3 (23.1%)	5 (38.5%)	13
TOTAL	15	38	24	148	225

Table 7.17: Frequency of VPAs in canonical and extraposited (extr.) subject position in the BNC

Canonical VP[to] subjects strongly favour *do so* over all others and also show a small preference for *do this* over *do that*. In extraposited VP[to] subjects, *do so* is again preferred, but *do it / this / that* are about as frequent. VP[ing] subjects are rare overall, and the frequencies reported here are too small for meaningful comparison.¹³

7.13 Other functions

This final section examines a range of cases where the function of the VPA sentence is unclear or not readily classifiable into one of the categories

12. Results for extraposited VP[ing] subjects were left out as a search with the same pattern as (2) returned just 4 occurrences, two of *do it* and two of *do that*.

13. COCA data for the same query show a strong preference for *do so* (380 occurrences), *do this* and *do that* (33 and 17 occurrences respectively); *do it* is again very rare (two occurrences)

defined above. All of the examples have the VPA in a *to*-infinitival clause except for one occurrence where it is a bare infinitive. The VPA clause may be either a non-finite relative in some of the VP[to] cases, or another type of infinitival subordinate. We will discuss these two subcategories in turn.

Examples of non-finite *to*-relatives include one which is comparable to the adjunct-choice cases of the type *a / the way to do it*: the VPA sentence states that a particular time is most suitable to carry out an action without specifying what that time is; this information is retrieved from the context.

(483) [W]hatever part of the tourist industry anybody's in erm now is the opportunity to benefit from erm the British Tourist Board's initiative and promote their own individual business erm featuring them in this erm massive advertising campaign, erm incorporating all the media, erm and benefit. They couldn't have a better time to do it. (KRL)

In two occurrences, the relative modifies an NP containing an ordinal, e.g., *the first ones in Britain to do it* (484). Strictly speaking, this type of structure, which does not allow VPE, specifies how many times the action has been performed previously, for instance, *we were the second ones to do it* means there has been one prior occurrence of the action, while *the first country to do it* implies there have been no previous occurrences. Discourse-wise, however, the purpose of the examples in (484)–(485) is more to emphasise that the action has been performed only once, or not at all.

(484) [Surrogate mother shares her experience]
 — And how did you actually, I mean, did you do all this er er er, as it were in an amateur way or did you do [it?] through any kind of profe, wi with the help of doctors or
 — We were, I think, the first ones in Britain to do it and it was all based on trust and if we made mistakes, we made mistakes! But

CHAPTER 7. FUNCTION OF THE VPA CLAUSE

we got through it in the end and they're happy, and I'm happy!
(FLG)

- (485) President Boris Yeltsin signed an agreement in July, er which effectively Russia recognise [sic.] Lithuania as an independent state, it was the second country in the world to do that, er all of this was sorted out in a very er, good, er peaceful, sensible way [...] (KJS)

The choice of *do it* in (484) is primarily a matter of saliency: in this case the antecedent is the main topic of conversation, and referring to it with *do all this* makes *do it* more likely than *do this / that* in subsequent mentions of the action. (485) allows *do it* and *do this* as well as *do that*; the latter is motivated both by the low saliency of the antecedent and its temporal remoteness.

(486) is similar in meaning to (483) above, but the VP[to] clause is not a relative, and is also not extraposed (cf. **To do that was time*). The nominal antecedent-trigger reduces the saliency of the antecedent, making *do it* less felicitous and VPE strongly dispreferred¹⁴. *Do this* would result in a focalisation through the subject's point of view, rather than Rita's.

- (486) And bit by bit their lovemaking turned into a dry ritual which caused Rita no actual pain, only a lingering, grey regret. Eventually she found other things to do when, in his view, it was time to do that. (AOR)

14. VPE is obviously rare in the idiomatic structure here independently of the context, but it is nevertheless infrequently attested: the following is an example from the COCA corpus (none were found in the BNC):

- (1) [NBC dateline news report *Quiet Storm; Georgia Power employees sue company for racial discrimination*; the black employee interviewed has just explained she was paid less than white colleagues in similar positions]
COURIC: (Voiceover) Wilson told us the company did give her a new title and a pay raise after she filed suit.
(Employees) Ms-WILSON: It's almost as a slap in the face. Why couldn't you just do it when it was time to to ? When the opportunity was there?

Two examples have the VPA in an AP modified by *enough* or *too*, functioning as the ‘indirect complement’ of the degree adverb, to use the term proposed by Huddleston and Pullum (2002, p. 443). *Do this* in (487) is motivated by the same focalisation effect which ruled it out in (486): the story is narrated through Nigel’s point of view, with a description of his opinions and state of mind. *Do it* and *do that*, while possible, would fail to achieve such an effect. (488) has a non-agentive, non-causal antecedent, preventing *do it*. *Do this* is extremely common in purpose adjuncts, as seen earlier in (7.9), but *do that* is also possible.

(487) Nigel began selling review copies—his own and any others he could cadge—to give a little boost to his income. He had been too lordly to do this at one time, preferring to give them as Christmas presents. (AC3)

(488) As the heavy stone had been placed in the middle of the circle, the ground had given way a little. It seemed to be made of rubber. The tape measure had now to go down into the hollow as well as across the circle, and it was not long enough to do this. (FNW)

The *to*-clause of (489) has essentially the same function as a conditional adjunct, allowing a paraphrase like *you’d be mad if you did it*. VPE is in theory possible (cf. *I’d be happy to*), but is prevented here by the distance from the antecedent-trigger, which is much too far back in the dialogue to give rise to a salient choice between alternatives. The imminent nature of the action favours *do this* over *do that* as an alternative; the latter would make the action appear more as a remote possibility, whereas the conversation makes it clear that Charles has every intention of joining the armed forces.

(489) [Charles is dining at a restaurant with his wife; he has just announced his intention to enlist in the army against his wife’s wishes.]
‘Men are going to be needed in the Services, mark my words.’

‘Yes, but they don’t mean you, darling.’

‘Why ever not? I’m an Englishman—an out-of-work one too. It seems I’m exactly what they mean.’

[The waiter arrives and talks with Charles, who requests a taxi to the barracks]

‘No, Charles!’ In her horror, Clarissa forgot all about the head waiter and the brandy. ‘Charles, you’d be mad to do it. It’s peace in our time, as Chamberlain said...’

‘Peace, hell. It’s war.’

(ACE)

The function of the VPA clause in (490) is much more straightforward than in previous examples: the point is simply to compare the antecedent action to that described in the previous clause. The comparative structure disallows VPE after the bare infinitive, but also in its *to*-infinitival variant (**Better to go on as before than to*).

(490) Nigel sometimes saw that he was a fool to have put up with everything. It was not something that he could admit to the outside world, though. Better to go on as before than do that. He told various stories when journalists interviewed him about the book and pryed [*sic.*] into his background. (AC3)

7.14 Conclusion

In this chapter we have identified a range of possible functions for the VPA clause. Results of the analysis suggest the single most common function of of VPAs is to describe additional properties of the antecedent action, what we have called ‘adjunct-choice’. Also common are complements, modals, and (to a lesser extent) VPE-type functions such as polarity or subject-choice (agent-choice, in our terminology). With respect to the latter, our data confirm the earlier suggestion made by Miller (2011) that

do it may occur in contexts otherwise licensing VPE of the antecedent is insufficiently salient, and show that this is also true of *do this/that*.

Beyond this general trend, there are some notable discrepancies between VPAs with respect to their typical discourse functions. Further description of an action is more commonly achieved with *do it* or *do this*; rarely *do that*. On the other hand, *do that* is favoured over *do it/this* in polarity-choice contexts, which is evidence of its greater ability to occur in contexts favouring VPE. Lastly, *do this* is rarely used to simply express a modal judgement on the action.

An obvious next step for future research is to analyse the possibility of alternation with VPE more systematically than has been done here. The proposed classification in its current form does not distinguish between cases where VPE is possible from those where it is not, yet it is clear that this information would be useful, since whether or not VPE is a valid alternative is of importance to the use of *do that*, which tends to be preferred over *do it/this* if VPE is acceptable. Determining the acceptability of VPE is not always simple, as there are many clear cases but also many others where judgements are less certain. What is certainly easier is to separate occurrences where VPE is grammatically possible and not systematically dispreferred—which includes all finite cases where there is no subsequent adjunct, and any non-finite ones where the VPA occurs after an auxiliary or in a *to*-complement—and those where it is ungrammatical (*to*-adjuncts and subjects, and V-ing forms) or strongly dispreferred (with adjuncts).

The classification established here was intended as exhaustive in an effort to capture the diversity of possible uses for VPAs. Systematic comparison with VPE would be greatly facilitated by a simpler annotation scheme, grouping together functions which may be subsumed under more general ones. A case in point is the function we have called ‘modal-choice’, which is ultimately a sub-case of polarity-choice, as are occurrences of finite VPAs in subordinate time clauses not followed by an adjunct.

Chapter 8

Conclusion

8.1 Summary of findings

Our study has clarified the discourse properties of *do it/this/that* on the basis of a larger sample of data than in previous corpus studies, showing that the alternation between them is non-trivial and depends on a number of factors, some of which can lead to a strong preference for one of the three, namely: register, the presence of a non-contrastive adjunct following the VPA, the discourse status of the antecedent at the point where it is introduced in discourse (i.e, whether it has been mentioned prior to the antecedent clause), the structure of the VPA clause and its discourse function, and the agentivity of the antecedent. To a lesser extent, VPA choice is also affected by the saliency of the antecedent, and its presumed familiarity to the addressee beyond the current discourse. Moreover, the choice of *do this* or *do that* is sensitive to the same criterion of speaker-inclusion which determines use of *this/that* in other contexts—whether or not the time/place of speaking is included in the time/place of the action. Based on these factors, we are now in a position to sketch a more precise picture of VPA alternation.

Do it is the preferred form to refer to highly salient, discourse-old an-

tecedents, especially those that are already established discourse topics and likely to remain so in subsequent discourse (a criterion which is also part [Gundel \(1993\)](#)'s in focus status). It is more typical in spoken than written register, and shows a somewhat stronger preference for agentive antecedents than *do this/that*. Its most typical use is what we have called 'adjunct-choice', specifying additional properties of the action by means of a non-contrastive adjunct. While this is also the most frequent use of *do this* and *do that*, it is in fact much more typical of *do it* (and atypical for *do that*). Exophorically, it is especially sensitive to the saliency of its antecedent, as it can only refer to actions that are ongoing in the speech situation, rather than being inferrable or made salient by it.

Do this, in contrast to *do it*, predominantly refers to actions that have not previously been mentioned prior to the antecedent clause, and thus constitute new information thus constitute new information in the antecedent clause. Often, *do this* introduces new information in the more general sense of facts previously unknown to the addressee. [Strauss \(2002\)](#)'s characterisation of *this* as signalling that the referent is important to the speaker and will be discussed further is readily applicable to *do this*, which is also commonly followed by an adjunct, or (sometimes extensive) further description of the antecedent by other means. These properties may explain its frequency in written discourse in contexts involving explanation or definition.

Do this is also the most frequent form of all in exophoric uses, which is proof that it requires a much less salient antecedent than *do it*; in particular, it allows reference to actions that are only inferrable from the extralinguistic context but otherwise not accessible to the addressee. The deictic use of *do this* exhibits all the properties of *this*: the antecedent action typically includes the speaker or is proximal to the speech situation, in particular by occurring within the speech situation, or shortly before or after it. In fiction narratives, *do this* may be used to achieve focalisation

through a particular character's point of view.

While it is also common in the adj-choice usage, it is not quite as frequent as *do it*, and when not followed by an adjunct, it occurs in certain contexts that are unusual for *do it*, such as subordinates expressing time or purpose. Statistically, it is more frequent in non-agentive uses than *do it/that*, but shows no real difference in acceptability with *do that*.

Do that differs from *do it/this* in preferring to occur without an adjunct. In this it is comparable to VPE, although the use of adjuncts after *do that* remains much more frequent, and does not systematically make it less acceptable. Further evidence of the similarity in usage between *do that* and VPE comes from its greater acceptability than *do it/this* in certain adjunct-less contexts typical of VPE, such as *if*-clauses, replies to polar questions, and, as shown in Chapter 7, a variety of polarity-selecting contexts where VPE may also be used, but *do it/this* are dispreferred. Miller and Hemforth (2014) argued that the point of using *do that* where VPE would be just as acceptable is to achieve certain discourse effects that would not obtain with VPE.

Analysis of the corpus data has provided a more precise characterisation of such effects, which ultimately have to do with the distal or speaker-exclusion sense of *that*: *do that* is apt to convey the idea that the action excludes the speaker (or is somehow 'far from' the speaker, in traditional terms), presenting it as unlikely, impossible, striking, or expressing some negative judgement about it (unpleasant, reprehensible, etc.). Though its most frequent function is adj-choice, it occurs in this function primarily when the adjunct occurs somewhere before the VPA (e.g., *Why did you do that/a way to do that*). Lastly, like *do this*, *do that* places lower requirements than *do it* on the saliency of its antecedent, and allows non-agentive uses more readily than *do it*.

8.2 Perspectives for future research

Our results are an important step in the study of a previously overlooked class of anaphoric expressions. However, it is in many respects simply a starting point for future research, which should uncover additional factors of alternation in addition to those examined. All aspects of VPA alternation considered here would obviously deserve further study in some way or other, but there are two areas in particular which seem to require attention.

8.3 Pending issues in VPA alternation

First, some important issues in the usage of VPAs have been insufficiently explained, or not at all, in the thesis. Due to the limitations of the BNC tagging and query syntax, it has not been possible to obtain precise data on the frequency of VPAs generally. It seems reasonable to assume that *do it* is more frequent than *do this/that*, it is difficult to predict how much more frequent it is. For the same reasons, the actual distribution of VPAs across registers remain unclear beyond the data reported in Chapter 2, which comprised a mix of fictional and actual dialogues, raising the question of how representative the fiction dialogues actually are.

8.3.1 Problems for the compositional analysis

A central tenet of the thesis has been the claim that VPAs are syntactically and semantically compositional, and as such inherit the properties of the object pronouns occurring after main verb *do*. Our results provide ample evidence of the analogy between pronoun choice in VPA and in other contexts with respect to saliency, discourse and information status, and the speaker inclusion/exclusion contrast with *do this/that*. In some

cases, however, variation between VPAs is less obviously attributable to the properties of the object pronouns. Agentivity is one such problem: there is a priori no reason to assume that having a demonstrative instead of *it* would make a non-agentive antecedent more acceptable. Likewise, detailed analysis of the types of adjuncts reveals interesting semantic tendencies, such as a preference for manner adjuncts with *do it*, and means adjuncts with *do this*. Again, it is unclear how these tendencies relate, if at all, to more general properties of *it* and *this*.

8.3.2 Experimental work

Our study is strictly corpus-based, and an experimental account of VPA choice remains to be done. Amongst the most relevant topics for experimental testing are adjuncts, agentivity, discourse status, and the most frequent functions identified in Chapter 7.

8.4 Comparison with other anaphoric expressions

The focus of the thesis was deliberately restricted to *do it/this/that*, although we have made some remarks on alternation with *do so* and VPE where relevant. A more systematic investigation of the alternation between VPAs and VPE is in order, especially seeing as there is evidence that the choice of *do it/this/that* can be influenced by whether or not VPE is possible. The analysis should also be extended to related anaphoric or elliptical constructions, such as *do the same*, *do likewise* as well as NCA or sluicing. One of [Tanenhaus and Carlson \(1990\)](#)'s experiments compared the processing of *do it* and NCA, but did not take into account any differences in usage between them.

Bibliography

- Aelbrecht, L. (2010). *The Syntactic Licensing of Ellipsis*. John Benjamins, Amsterdam.
- Ariel, M. (1996). [Linguistic Marking of Physical Givenness](#). Paper presented at the 2nd International Conference on Deixis: ‘Time, Space and Identity’, CRIN, Nancy, 28-30 Mar.
- Arregui, A., Clifton Jr, C., Frazier, L., and Moulton, K. (2006). [Processing elided verb phrases with flawed antecedents: The recycling hypothesis](#). *Journal of Memory and Language*, 55(2):232–246.
- Atkinson, J. and Heritage, J. (1994). *Structures of Social Action: Studies in Conversation Analysis*. Studies in emotion and social interaction. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Grammar of spoken and written English*. Harlow: Longman.
- Bolinger, D. (1970). [The meaning of do so](#). *Linguistic Inquiry*, 1(1):140–144.
- Borthen, K., Fretheim, T., and Gundel, J. K. (1997). [What brings a higher-order entity into focus of attention? Sentential pronouns in English and Norwegian](#). In *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, pages 88–93. Association for Computational Linguistics.

BIBLIOGRAPHY

- Bouton, L. (1970). Do so: Do+ adverb. In Sadock, J. and Vanek, A., editors, *Studies Presented to Robert B. Lees by his Students*, pages 17–38. Edmonton: Linguistics Research.
- Cohen, A. (1980). [On the graphical display of the significant components in two-way contingency tables](#). *Communications in Statistics - Theory and Methods*, 9(10):1025–1041.
- Cornish, F. (1992). [So Be It: The Discourse–Semantic Roles of so and it](#). *Journal of Semantics*, 9(2):163–178.
- Cornish, F. (1996). [‘Antecedentless’ anaphors: deixis, anaphora, or what? Some evidence from English and French](#). *Journal of Linguistics*, 32(1):19–41.
- Cornish, F. (1999). *Anaphora, discourse, and understanding: evidence from English and French*. Oxford University Press.
- Cruse, D. A. (1973). [Some Thoughts on Agentivity](#). *Journal of Linguistics*, 9(1):11–23.
- Culicover, P. W. and Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.
- Dalrymple, M., Shieber, S. M., and Pereira, F. C. (1991). [Ellipsis and higher-order unification](#). *Linguistics and philosophy*, 14(4):399–452.
- Davies, M. (2004). [BYU-BNC](#). based on the british national corpus from oxford university press.
- Davies, M. (2008). [The Corpus of Contemporary American English \(COCA\): 560 million words, 1990–present](#). Available online at <https://corpus.byu.edu/coca/>.
- Desagulier, G. (2017). *Corpus Linguistics and Statistics with R*. Springer.

- Dowty, D. (1991). [Thematic Proto-Roles and Argument Selection](#). *Language*, 67(3):547–619.
- Dubey, A., Sturt, P., and Keller, F. (2005). [Parallelism in Coordination As an Instance of Syntactic Priming: Evidence from Corpus-based Modeling](#). In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, pages 827–834, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Fillmore, C. (1968). [The Case for Case](#). In Bach, E. and Harms, R., editors, *Universals in Linguistic Theory*. New York: Holt, Rinehart, & Winston.
- Fillmore, C. (1982). Towards a descriptive framework for spatial deixis. In Jarvella, R. J. and Klein, W., editors, *Speech, Place, and Action Studies in Deixis and Related Topics*. New York: John Wiley & Sons.
- Frazier, L. and Clifton, Jr., C. (2001). [Parsing Coordinates and Ellipsis: Copy \$\alpha\$](#) . *Syntax*, 4(1):1–22.
- Frazier, L., Munn, A., and Clifton, C. (2000). [Processing Coordinate Structures](#). *Journal of Psycholinguistic Research*, 29(4):343–370.
- Frazier, L., Taft, L., Roeper, T., Clifton, C., and Ehrlich, K. (1984). [Parallel structure: A source of facilitation in sentence comprehension](#). *Memory & Cognition*, 12(5):421–430.
- Friendly, M. (2000). *Visualizing categorical data*. SAS Institute, Cary, NC.
- Garnham, A. and Oakhill, J. (1987). [Interpreting elliptical verb phrases](#). *The Quarterly Journal of Experimental Psychology*, 39(4):611–627.
- Gruber, J. S. (1967). [Look and see](#). *Language*, 43(4):937–947.
- Gundel, J. K. (1993). [Cognitive Status and the Form of Referring Expressions in Discourse](#). *Language*, 69(2):274–307.

BIBLIOGRAPHY

- Gundel, J. K., Borthen, K., and Fretheim, T. (1999). The role of context in pronominal reference to higher order entities in English and Norwegian. In *International and Interdisciplinary Conference on Modeling and Using Context*, pages 475–478. Springer.
- Halliday, M. A. K. (1967). [Notes on Transitivity and Theme in English: Part 1](#). *Journal of Linguistics*, 3(1):37–81.
- Hankamer, J. and Sag, I. (1976). [Deep and Surface Anaphora](#). *Linguistic Inquiry*, 7(3):391–428.
- Hardt, D. (1993). [Verb Phrase Ellipsis: Form, Meaning, and Processing](#). PhD thesis, University of Pennsylvania.
- Hegarty, M. (2003). [Verb Phrase Ellipsis: Form, Meaning, and Processing](#). *Lingua*, 113(9):891–927.
- Higgins, F. (1992). Vicarious *do* and the auxiliary. Ms., University of Massachusetts, Amherst.
- Houser, M. J. (2010). [The Syntax and Semantics of Do So Anaphora](#). PhD thesis, University of California, Berkeley.
- Huddleston, R. and Pullum, G. (2002). *The Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Kehler, A. (2000). [Coherence and the resolution of ellipsis](#). *Linguistics and Philosophy*, 23(6):533–575.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications.
- Kehler, A. and Ward, G. (1999). [On the semantics and pragmatics of identifier so](#). In Turner, K., editor, *The Semantics/Pragmatics Interface from Different Points of View (Current Research in the Semantics/Pragmatics Interface Series, Volume I)*. Amsterdam: Elsevier.

- Kertz, L. (2008). [Focus structure and acceptability in verb phrase ellipsis](#). In Abner, N. and Bishop, J., editors, *Proceedings of the 27th West Coast Conference on formal linguistics*, pages 283–291.
- Kertz, L. (2010). [Ellipsis reconsidered](#). PhD thesis, University of California at San Diego, Unpublished PhD dissertation.
- Kertz, L. (2013). [Verb phrase ellipsis: The view from information structure](#). *Language*, 89:390–428.
- Lakoff, G. and Ross, J. R. (1976). Why you can't do so into the sink. In McCawley, J. D., editor, *Syntax and Semantics, Volume 7: Notes from the Linguistic Underground*. Academic Press.
- Levine, R. (2012). [Auxiliaries: To's company](#). *Journal of Linguistics*, 48:187–203.
- Lowder, M. and Gordon, P. C. (2015). [Natural Forces as Agents: Reconceptualizing the Animate-Inanimate Distinction](#). *Cognition*, 136C:85–90.
- Maurer, G., Tanenhaus, M. K., and Carlson, G. N. (1995). [A note on parallelism effects in processing deep and surface verb-phrase anaphora](#). *Language and Cognitive Processes*, 10(1):1–12.
- Merchant, J. (2004). [Fragments and ellipsis](#). *Linguistics and Philosophy*, 27(6):661–738.
- Michiels, A. (1978). [A Note on the Relation between Agent and Stativity](#). *Neophilologus Groningen*, 62(2)(62.2):172–177.
- Miller, P. (2002). [Les emplois non finis de do auxiliaire](#). In Delmas, C., editor, *Construire et reconstruire en linguistique anglaise: syntaxe et sémantique*, pages 185–198. Publications de l'Université de Saint-Etienne, Saint-Etienne.

BIBLIOGRAPHY

- Miller, P. (2011). [The Choice between Verbal Anaphors in Discourse](#). In Hendrickx, I., Devi, S. L., Branco, A., and Mitkov, R., editors, *Anaphora Processing and Applications: 8th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2011, Faro Portugal, October 6-7, 2011. Revised Selected Papers*. Springer.
- Miller, P. (2013). [Usage Preferences: The Case of the English Verbal Anaphor *do so*](#). In Müller, S., editor, *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, HPSG Online Proceedings.
- Miller, P. and Hemforth, B. (2014). [Verb Phrase Ellipsis with Nominal Antecedents](#). Ms., Université Paris Diderot (submitted).
- Miller, P. and Pullum, G. K. (2014). [Exophoric VP ellipsis](#). In Hofmeister, P. and Norcliffe, E., editors, *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, pages 5–32.
- Murphy, G. L. (1985). [Processes of understanding anaphora](#). *Journal of Memory and Language*, 24(3):290–303.
- Murphy, G. L. (1990). [Interpretation of verb phrase anaphora: Influences of task and syntactic context](#). *The Quarterly Journal of Experimental Psychology Section A*, 42(4):675–692.
- Méry, R. (2002). [A propos du verbe DO](#). *Anglophonia*, 12:81–113.
- Phillips, C. and Parker, D. (2014). [The psycholinguistics of ellipsis](#). *Lingua*, 151:78–95.
- Prince, E. (1981). Toward a taxonomy of given-new information. In Cole, P., editor, *Radical pragmatics*. Academic Press.

- Prince, E. F. (1986). [On the syntactic marking of presupposed open propositions](#). In *Proceedings of the 22nd Annual Meeting of the Chicago Linguistic Society*.
- Pullum, G. K. (1991). [English nominal gerund phrases as noun phrases with verb phrase heads](#). *Linguistics*, 29:763–799.
- Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., and Crystal, D. (1985). *A comprehensive grammar of the English language*. Cambridge University Press.
- Roberts, L., Matsuo, A., and Duffield, N. (2013). [Processing VP-ellipsis and VP-anaphora with structurally parallel and nonparallel antecedents: An eye-tracking study](#). *Language and Cognitive Processes*, 28(1-2):29–47.
- Ross, J. (1970). Act. In Davidson, D. and Harman, G., editors, *Semantics of Natural Language*, volume 40 of *Synthese Library*, pages 70–126. Springer Netherlands.
- Sag, I. (1976). [Deletion and Logical Form](#). PhD thesis, Massachusetts Institute of Technology.
- Sag, I. A. and Hankamer, J. (1984). [Toward a Theory of Anaphoric Processing](#). *Linguistics and Philosophy*, 7(3):325–345.
- Schlesinger, I. M. (1989). [Instruments as Agents: On the Nature of Semantic Relations](#). *Journal of Linguistics*, 25(1):189–210.
- Schuster, E. (1988). Pronominal reference to events and actions: evidence from naturally-occurring data. Technical report, University of Pennsylvania.
- Sharifzadeh, S. (2018). L’anaphore prédicative en anglais : étude comparative des reprises en DO SO, DO IT, DO THIS, DO THAT et DO opérateur. In *Marqueurs et Structures*. Lambert-Lucas.

BIBLIOGRAPHY

- Shopen, T. (1972). *A generative theory of ellipsis*. PhD thesis, UCLA.
- Simner, J., Pickering, M., and Garnham, A. (2003). [Discourse Cues to Ambiguity Resolution: Evidence From “Do It” Comprehension](#). *Discourse Processes*, 36(1):1–17.
- Simner, J. C. (2001). *Engaging long and short term memory during anaphor comprehension*. PhD thesis, University of Sussex.
- Sobin, N. (2008). [Do so and VP](#). *Linguistic Inquiry*, 39(1):147–160.
- Souesme, J.-C. (1985). *Do something et ses diverses réalisations en anglais contemporain*. PhD thesis, Université de Nice.
- Strauss, S. (2002). [this, that, and it in spoken american english: a demonstrative system of gradient focus](#). *Language Sciences*, 24:131–152.
- Stroik, T. (2001). [On the light verb hypothesis](#). *Linguistic Inquiry*, 32(2):362–369.
- Tanenhaus, M. K. and Carlson, G. N. (1990). [On the light verb hypothesis](#). *Language and Cognitive Processes*, 5(4):257–280.
- Ward, G. and Kehler, A. (2005). [Syntactic form and discourse accessibility](#). In Branco, A., McEnery, T., and Mitkov, R., editors, *Anaphora processing: Linguistic, cognitive and computational modelling*, pages 365–384. Amsterdam: Benjamins.
- Zeileis, A., Meyer, D., and Hornik, K. (2007). [Residual-based shadings for visualizing \(conditional\) independence](#). *Journal of Computational and Graphical Statistics*, 16(3):507–525.