

Lexical variation of expressions for ‘DIE’: A mixed methods inquiry

by

Laura Ford

A Thesis submitted to the Faculty of Graduate and Postdoctoral Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Linguistics

University of Manitoba

Winnipeg

Copyright © 2026 by Laura Ford

Except as otherwise permitted by any open licensing applied to this work, the author expressly reserves all rights to use of this work for Generative Artificial Intelligence training, the development of machine learning language models, and similar technologies. Where open licensing applies to this work, the author expresses their preference to opt out of the preceding uses.

Examining Committee

Supervisor:

Dr. Verónica Loureiro-Rodríguez

Department of Linguistics

University of Manitoba

Members:

Dr. Terry Janzen

Department of Linguistics

University of Manitoba

Dr. Kevin Russell

Department of Linguistics

University of Manitoba

Abstract

This thesis uses mixed-methods to investigate variation in the lexical variable ‘DIE,’ focusing on the variants *DIE*, *PASS away*, *PASS on* and *PASS*. First, taking a qualitative approach, it analyzes attitudes towards variation expressed in a corpus of Reddit commentary made up of four threads and 254 comments. Using Thematic Analysis and a language attitudes framework, I identify and explore four themes centred on variation of ‘DIE’: nuances of variant meaning, the relevance of emotions, communitive context, and associations with macrosocial categories. Next, guided by the results of the qualitative analysis, I follow quantitative variationist methods to analyze ‘DIE’ as a sociolinguistic variable. To capture the variable in use, data was collected from the spoken portions of the Canadian Business and Current Affairs database and the Corpus of Contemporary American English, coding 1000 tokens for linguistic and social factors that may influence variant choice. Multiple statistical methods including random forest analysis, conditional inference tree analysis, and logistic regression modeling allow exploration of the influence and interrelationships of many potential predictors. Of these, DEATH TYPE, PROGRESSIVE, DEATH HAS OCCURRED and PERSONAL RELATIONSHIP stand out. Violent or sudden death, deaths that have not occurred and possibly use of the progressive favoured the direct language variant *DIE*, while contexts where the speaker had a personal relationship with the deceased favour use of the euphemistic variants *PASS away*, *PASS on* and *PASS*. Both the qualitative and quantitative studies highlight the importance of the metaphorical nature of PASS-type variants to factors that condition variation. Together, the two studies characterize *DIE* as informative, bodily focused, and impersonal and *PASS away*, *PASS on* and *PASS* as specialized variants recruited principally for relationship-focused communication.

Keywords: lexical variation, ‘DIE’, Canadian English, mixed methods, corpora

Acknowledgements

I’d like to thank the University of Manitoba and Faculty of Graduate and Post-doctoral Studies for their financial support through the University of Manitoba Graduate Fellowship and Faculty of Graduate Studies Research Completion Scholarship. This funding has made a world of difference to me and to my family.

I cannot thank my advisor, Dr. Loureiro-Rodríguez, enough for her guidance, invaluable feedback and bottomless patience. She encouraged, questioned and pushed me exactly when I needed it and, without question, this project would not have happened without her. Thank you to the members of my examining committee for their insight and expertise, and to Dr. Russell for the mountains of indispensable help with the statistical analysis. To the Consistent Linguists™ (you know who you are!), thank you for the accountability, encouragement and stress-relieving laughs.

To my family, your support and belief in me is my solid ground.

Table of Contents

List of Tables	viii
List of Figures	ix
List of Typographical Conventions	x
Chapter 1: Introduction	1
Chapter 2: “Why did the queen “die peacefully”?” a corpus-based attitudinal study	3
2.1 Introduction.....	3
2.2 Background.....	3
2.3 Data and methods.....	5
2.3.1 Corpus	5
2.3.2 Thematic analysis.....	7
2.4 Nuances of variant meaning.....	10
2.5 Relevance of emotions.....	13
2.6 Relevance of communicative context	16
2.7 Associations with macrosocial categories: regional variation, social class.....	21
2.8 Conclusion	23
Chapter 3: “At some point there’s nobody there and the individual passes”: a corpus-based study of lexical variants of ‘DIE’	26
3.1 Introduction.....	26
3.2 Corpora in sociolinguistic research.....	26
3.3 Lexical variation	31
3.4 ‘DIE’ as a linguistic variable	34
3.5 Methodology	38
3.5.1 Corpus data.....	39

3.5.2 Circumscribing the variable context	40
3.5.2.1 Corpus queries	40
3.5.2.2 Exclusion criteria	41
3.5.3 Coding	45
3.5.3.1 Linguistic factors	45
3.5.3.2 Social factors	45
3.5.4 Statistical analysis	48
3.6 Results	51
3.6.1 Descriptive statistics	51
3.6.2 Random forest analysis	52
3.6.3 Conditional inference tree analysis	55
3.6.4 Logistic regression	57
3.7 Discussion	60
3.7.1 With what relative frequency do variants of ‘DIE’ occur in spontaneous media discourse?	60
3.7.2 Does use of variants differ according to social factors?	61
3.7.3 Does use of variants in this context differ according to linguistic factors?	66
3.7.4 Does use of variants differ between Canada and the United States?	69
3.8 Conclusion	69
Chapter 4: Conclusion	73
References	75
Appendix A: Frequency of direct and euphemistic variants of ‘DIE’ in Southeastern United States (Johnson, 1996, p. 272)	88

Appendix B: Frequency of <i>PASS</i> -type variants of ‘DIE’ in in Southeastern United States (Johnson 1996, p. 272)	88
--	----

List of Tables

Table 1 Sources of corpus data for Thematic Analysis	7
Table 2 Initial deductive codes for Thematic Analysis	8
Table 3 Results of Thematic Analysis	9
Table 4 Summary of nuances of variant meaning	12
Table 5 Rates of direct language use in linguistic and medical literature	35
Table 6 Corpus data collection queries.....	41
Table 7 Inclusion/exclusion criteria and their application.....	42
Table 8 Best-fit model predicting <i>DIE</i> (versus <i>PASS</i> -type variants, N=528) compared to models with one fewer, one more and one alternative predictor	58
Table 9 Overall best-fit model predicting <i>DIE</i> (versus <i>PASS</i> -type variants, N=528).....	59

List of Figures

Figure 1 Affective stances toward variants of ‘DIE’ expressed by commenters (n=77).....	13
Figure 2 Reported use of variants of ‘DIE’ by commenter (n=121)	16
Figure 3 Pragmalinguistic versus sociopragmatic politeness (Leech, 2014, p. 17-18, Figs. 1.1-2)	21
Figure 4 Circumscribing the variable context of a lexical variable (variants in green).....	32
Figure 5 Overall distribution of variants of ‘DIE’ (N=1000).....	51
Figure 6 Distribution of variants of ‘DIE’ by country (N=1000).....	52
Figure 7 Conditional permutation importance for the choice between <i>DIE</i> and <i>PASS</i> -type variants, all potential predictors.....	53
Figure 8 Example conditional inference trees from the random forest modeling the choice between <i>DIE</i> and <i>PASS</i> -type variants.....	54
Figure 9 Conditional inference tree for the choice between <i>DIE</i> and <i>PASS</i> -type variants.....	56
Figure 10 Conditional inference tree for the choice between <i>PASS away</i> , <i>PASS on</i> and <i>PASS</i>	57

List of Typographical Conventions

linguistic form.....	italics.....	<i>died</i>
lemma of a linguistic form.....	italics, small capitals.....	<i>DIE</i>
lemma of a linguistic variable.....	single quotation marks, small capitals.....	‘DIE’
query or search term.....	double quotation marks, as typed.....	“DIE_v”
predictor variable.....	small capitals.....	DEATH TYPE

Chapter 1: Introduction

When an English-speaker needs to communicate that someone has died, they have many options: *died* itself, one of many euphemisms such as *passed away*, *passed*, *left us* or *went home to the Lord*, or even a dysphemism like *croaked*. These may be thought of as lexical variants of the variable ‘DIE.’ Recent high-profile initiatives like a report commissioned by *The Lancet* to challenge the alienation of families and communities from medicalized death systems (Sallnow et al., 2022) indicate that societal engagement with death in North America is potentially in flux and certainly a concern of the moment. Though use of these variants has been studied by other disciplines, it has rarely been undertaken from a linguistic point of view, and little is known about differences in the meanings and uses of these variants.

This thesis investigates variation in ‘DIE’ to contribute to the study of lexical variation and gain insight into how language use reflects societal perceptions of a universal truth. To do so I use three corpora and two complementary methodologies. First, Chapter 2 takes a qualitative approach, analyzing attitudes towards variation expressed in a corpus of Reddit commentary made up of four threads and 254 comments. Using Thematic Analysis (Braun & Clarke, 2006) and a language attitudes framework, I identify and explore four themes centred on variation of ‘DIE’: nuances of variant meaning, the relevance of emotions, communitive context, and associations with macrosocial categories.

Guided by the results of this qualitative analysis, Chapter 3 uses quantitative variationist methods to analyze ‘DIE’ as a sociolinguistic variable. To capture the variable in use, I collect data from the spoken portions of the Canadian Business and Current Affairs database and the Corpus of Contemporary American English, coding 1000 tokens for linguistic and social factors that may influence variant choice. Multiple statistical methods including random forest analysis,

conditional inference tree analysis, and logistic regression modeling allow me to explore the influence and interrelationships of many potential predictors. Of these, DEATH TYPE, DEATH HAS OCCURRED, PERSONAL RELATIONSHIP and PROGRESSIVE, stand out. Violent or sudden death, deaths that have not occurred and possibly use of the progressive favoured the direct language variant *DIE*, while contexts where the speaker had a personal relationship with the deceased favour use of the euphemistic variants *PASS away*, *PASS on* and *PASS*.

Chapter 4 synthesizes findings from both studies and concludes the thesis.

Chapter 2: “Why did the queen “die peacefully”?” a corpus-based attitudinal study

2.1 Introduction

The goal of this chapter is to investigate attitudes toward variants of ‘DIE’ using a language attitudes framework, which considers cognitive, affective, and behavioural attitudinal components. I take a societal treatment approach (Kircher, 2022), using Thematic Analysis (Braun & Clarke, 2006) to examine preexisting commentary on the topic readily available on the social question and answer (social Q&A) platform Reddit in order to answer the following research question:

Regarding variants of ‘DIE’, what do commenters on social media say they believe, feel and use?

To answer this question, I begin by summarizing what we know about variants of ‘DIE,’ largely from fields outside of linguistics (2.2). Next, section 2.3 describes the corpus of social media commentary studied and methods employed. Sections 2.4-2.7 discuss four major themes identified in the corpus: nuances of variant meaning, relevance of emotions, relevance of communicative context, and associations with macrosocial categories. Section 2.8 concludes.

2.2 Background

Variants of ‘DIE’ have been studied in two main contexts, social science studies of obituaries and epitaphs, and in the medical literature, where both clinician and patient perspectives are examined.

In the past, direct, non-euphemistic language was used more often in North American obituaries than today (Ford & Loureiro-Rodríguez, forthcoming; Phillips, 2007). Traditional views of death saw it as an inescapable natural phenomenon and it was spoken about frankly, before a gradual shift toward euphemistic language, as it became appropriate to resist death via

trust in medical science. At present, the euphemistic variant *PASS away* is the dominant variant in central Canada and the US, accounting for approximately 75% of use in obituaries in both Winnipeg, MB and Grand Forks ND (Ferraro, 2019; Ford & Loureiro-Rodríguez, forthcoming), but was rarely seen in obituaries or epitaphs in the early twentieth century (Smithson, 2019). Since that time, variation in expressions for ‘DIE’ in obituaries has increased, and writers have many choices. Characteristics of the obituary writer and the deceased—who can be thought of as participants in the obituary—can impact choice of variant. Higher religiosity favours use of other euphemisms such as *left us* and *went to be with the Lord* over *passed away* (Ford & Loureiro-Rodríguez, forthcoming). The role of gender is less clear. Ferraro (2019) reports that female gender of the deceased favours use of *PASS away* over *DIE* while other investigations observe no effect (Ford & Loureiro-Rodríguez, forthcoming; Matiki, 2001). *PASS*, a variant that only appears after the turn of the century in central Canadian obituaries, may be less formal than *PASS away* (Ford & Loureiro-Rodríguez, forthcoming).

Use of *PASS* and *PASS away* is also common in medical practice, both when discussing the possibility of patient death with families, and when informing colleagues in writing that a death has occurred in hospital (Barlet et al., 2022; Habeck-Fardy, 2021). Euphemisms based on medical terminology (e.g. *irrecoverable drop in heart rate*, metonymic use of *code* or *stop breathing*) were used even more frequently than *PASS away* by clinicians, despite official recommendations to use non-euphemistic direct language in order to avoid misunderstandings.

In addition to examining clinician behaviour, the medical research community has undertaken qualitative study of the general population to better understand patient preferences and attitudes toward discussing death, including use of variants of ‘DIE’ (Rawlings et al., 2017; E. Wilson et al., 2024). Participants in these studies felt that less contact with death today may make

people in contemporary society more reluctant to talk about it directly. They also characterized their own use of euphemisms as context-dependant, varying with setting and interlocutor.

Euphemisms were considered ‘soft,’ sensitive to others, and less likely to cause distress or pain by many, though some rejected the notion that they are able to lessen grief. Many said they used direct language themselves, but understood why others would choose euphemisms. They also identified euphemisms, particularly *left us* and *passed*, as a potential source of misunderstanding.

Overall, available research suggests euphemistic variants of ‘DIE’ are commonly used in multiple contexts in contemporary society, and that among them *PASS*-type euphemisms are especially frequent. Among medical professionals it appears that practice guidelines (cognition) and use in practice (behaviour) are not always aligned. We might infer that, compared to direct language, use of euphemisms is evaluated more positively by the general population, considering their characterization as sensitive to others, but this is far from definitive. Although choice of variant of ‘DIE’ is described as context dependant, which contexts favour which variants has not been explored. Considering this, much remains to be clarified regarding cognitive, affective and behavioural components of attitudes toward variants of ‘DIE.’

2.3 Data and methods

2.3.1 Corpus

Identification of data for analysis began with the internet search query “Died, passed or passed away?” which yielded hits from the social Q&A platforms Reddit and Quora. As a greater volume of relevant commentary was available on Reddit (22 discrete threads versus fewer than ten overlapping threads), it was chosen as the data source for this inquiry. Four threads were selected according to the following inclusion criteria. Individual threads were to be as recent as possible, have non-confrontational original posts and fewer than 300 comments so that all

comments could be analyzed. Non-confrontational posts were preferred, as commenters spent more time on topic in these threads, and less time on off-topic insults and dismissals. As a group, the threads were to discuss the three variants of greatest interest based on the literature review (*DIE*, *PASS away*, and *PASS*) as well as use in a variety of contexts. They were also chosen to balance original posts which questioned use of non-euphemisms with those that questioned use of euphemisms according to comment volume.

Details of the four threads and associated comments which make up the corpus are given in Table 1. Before the application of comment exclusion criteria these constituted 366 comments. Any comment that was deleted, redacted (replaced with a set of nonsense words by a redaction program), authored by a moderator bot, or solely expressed the poster's thanks was excluded. Comments regarding expressions for specific types of death were excluded (e.g. drowned, was murdered), as were comments deemed too general (e.g. discussion of non-death euphemisms or expressions for 'DIE' in languages other than English) as well as those that veered completely off topic (e.g. discussion of whether policing is a respectable profession). After application of exclusion criteria 254 comments remained.

Table 1*Sources of corpus data for Thematic Analysis*

Date	Thread	Comments retained (total)	Perspective/context	% Corpus
1 Apr. 2024	Do you say a person died or passed away?	145 (234)	open query	57%
www.reddit.com/r/questions/comments/1btoavt/do_you_say_a_person_died_or_passed_away/			professional/ non-professional	
8 Sep. 2022	Why did the queen "die" peacefully and not "pass away" peacefully? Doesn't "died" sound a bit harsh?	58 (71)	questions use of <i>DIE</i> non-professional	23%
www.reddit.com/r/ENGLISH/comments/x9l3a1/why_did_the_queen_die_peacefully_and_not_pass				
3 Dec. 2024	"Passed Away, Passed On, Passed?"	27 (37)	questions use of <i>PASS</i> euphemisms	20%
http://www.reddit.com/r/questions/comments/1h60frg/passed_away_passed_on_passed/				
2 May 2024	Passed instead of died	24 (24)	non-professional	
www.reddit.com/r/ENGLISH/comments/1cilowk/passed_instead_of_died/				
		254 (366)		100%

2.3.2 Thematic analysis

Retained comments were saved in a word processing document and imported to the qualitative analysis program Delve. Here the data was coded recursively and analyzed using Thematic Analysis (Braun & Clarke, 2006). This process began with a set of deductive codes (Table 2): 1) semantic codes to track which variants commenters discussed, 2) language attitude framework codes (cognitive, affective, behavioural) and 3) codes otherwise suggested by the literature.

Table 2*Initial deductive codes for Thematic Analysis*

Semantic codes	non-euphemism, dysphemism, PASS away, PASS on, PASS, PASS euphemisms in general, euphemisms in general
Language attitudes framework codes	cognitive, affective, behavioural
Codes suggested by the literature	religion, change over time, formality, misunderstanding/clarity, gender, harshness, softness, grief, fear, avoidance, taboo, setting, interlocutor, professional context, personal context, talking to children

Semantic coding confirmed that *PASS*-type variants are the most frequently discussed euphemisms in the corpus accounting for 67% of euphemism mentions. Language attitudes codes were quickly abandoned, as those suggested by the literature are largely cognitive and affective in nature, but far more detailed. The ‘behavioural’ code was also made more specific and used to track ‘reported use.’ The only deductive code suggested by the literature which was not employed or adapted was gender, as this was not discussed at all in the corpus. These, along with codes generated inductively during the recursive coding process resulted in approximately sixty total codes, which eventually formed the four themes and twelve subthemes given in Table 3.

Table 3
Results of Thematic Analysis

Theme	Subtheme	Details
nuances of variant meaning	shared denotative meaning	clarity, honesty is proportional to accessibility of denotative meaning
	different connotative meanings	afterlife beliefs
		manner of death
	different affective meanings	harshness
		softness
relevance of emotions	affective stances toward variants	+ve toward non-euphemism
		-ve toward non-euphemism
		+ve toward euphemism
		-ve toward euphemism
	experience of grief	euphemism helps
		euphemism doesn’t help
	discomfort strategies: avoidance and humour	entertainment media, dysphemism
relevance of communicative context	informative communication favours non-euphemisms	journalism
		informing next of kin
	supportive communication favours euphemisms	grieving interlocutor
		what others might want/find comforting
	social distance	from decedent, between interlocutors, standard vs. non-standard variants
	(in)formality and (im)politeness	
associations with macrosocial categories	regional variation	
	variation by social class	

The first major theme, nuances of variant meaning, relates to the denotative, connotative and affective meanings of variants. This includes discussion of whether variants are equivalent or non-equivalent and how they came to mean what they do. The second theme, relevance of emotion, involves how people feel about different variants, how variants relate to the experience and processing of grief, and the emotional strategies of avoidance and humour. Theme three, the relevance of social context, considers the importance of interactional factors to variant choice including communicative purpose, and social distance between interactional participants. The final theme, associations with macrosocial categories, focuses on the subthemes of regional variation and variation by social class.

2.4 Nuances of variant meaning

Regardless of commenter preferences for variants of 'DIE' there was a general acknowledgement that these variants denote the same event, either by saying so explicitly, as in example (1) or more indirectly with phrases like "dead is dead" in (2).

- (1) How is saying, "I'm sorry your husband died" crude compared to saying, "I'm sorry your husband passed away"? You're still saying the exact same thing.
- (2) I say passed away,I don't know why,it just seems softer,either way dead is dead

The accessibility of that denotative meaning, however, differed between variants. Commenters framed this in two ways: clarity and honesty. Non-euphemistic variants were characterized as clear and factual (3). In contrast, the potential for euphemisms to lead to misunderstanding due to lack of clarity was highlighted (4), and they were also seen by some as willfully if partially obscuring the truth of death (5).

- (3) That's somebody died does not sound harsh at all to my ears but rather just objectively recording the fact.

(4) I learned that you need to be direct & not use euphemisms many years ago. A person died and the doctor went out to tell them. I dont knoe what he said. But next thing they are coming in? And suddenly a woman (his wife i think) starts screaming "oh my god! Hes dead! He's DEAD!" [...] Causr whatever the doctor said to them? They didn't realise the person was dead.

(5) Perhaps you cannot face the reality of death and have to hide behind baby words but I have seen most of my family die and they died - they didnt "pass away".

In addition to a shared denotative meaning, variants can have different connotative meanings, particularly regarding afterlife beliefs and manner of death. Commenters felt that *PASS*-type euphemisms are religious in origin (6) and carry the connotation that the deceased continues to exist after death. This continued existence need not be dogmatically religious however, as hinted in (6; "whatever that may be") and more fully expressed in (7). *DIE*, in contrast, was characterized as final, and did not carry afterlife connotations (8).

(6) I think it stems from religion..."they've passed" meaning they've passed from this life to the next, whatever that may be.

(7) [Passed] sort of implies that she has gone on to somewhere else. Perhaps a better place, I hope. Perhaps to a wonderful journey. Perhaps just to some place where she might rest peacefully.

(8) Died. They don't pass "to" anywhere. They no longer exist. Sorry.

Euphemistic variants and non-euphemistic variants, typically exemplified by *PASS away* and *DIE* respectively, also carried a different set of dichotomous connotative meanings related to manner of death. Many commenters felt that *PASS away* suggested a death related to advanced

age or due to a slowly progressing illness, but that *DIE* implied a sudden and unexpected death.

Example (9) is representative of many comments of this nature.

(9) Depends... Idk why but in my head, passed away = natural causes/illness/age

Died = anything else/accident/unexpected

Although there were some who disagreed, as in example (3) above, non-euphemistic variants were typically characterized as harsh, as in (10), while euphemistic variants were characterized as soft (11), which was consistent with previous literature (Rawlings et al., 2017).

(10) I feel "died" is very harsh and abrupt, kind of unfeeling.

(11) I think I prefer 'passed'. Its softer, like my thoughts of her are always soft and fond.

Nuances of variant meaning appear to constitute a set of opposites with one polar value related to non-euphemisms and the other to euphemisms. These are summarized in Table 4. Initially it may seem that aspects of meaning as diverse as accessibility of denotative meaning and manner of death could not be related, nevertheless there is thematic consistency across categories. It is possible that there is an underlying aspectual contrast that creates these opposites, which I have expressed visually in the final row of Table 4.

Table 4

Summary of nuances of variant meaning

Type of meaning	Non-euphemism	Euphemism
denotative	clear	vague
connotative (conception of death)	final	non-final
connotative (manner of death)	abrupt	gradual
affective	harsh	soft
visual representation ¹	●	●

¹ This is the author's subjective impression and does not have a theoretical underpinning.

The nature and implications of this aspectual contrast will be explored further in Chapter 3 sections 3.7.2 and 3.7.3.

2.5 Relevance of emotions

Commenters took various affective stances toward a variety of variants, thus coding was simplified to positive or negative affective stance toward euphemism or non-euphemism. The frequency of responses that express these four possibilities is illustrated in Figure 1, where we see that far more commenters took a negative affective stance toward euphemisms than any other possibility represented. These commenters expressed their feelings in a number of ways. Many simply said they disliked euphemisms as in (12) while others characterized them as illogical (13) or ridiculously vague (14). Those who took exception to euphemisms, however, did not necessarily have a problem with dysphemisms as we see in (13).

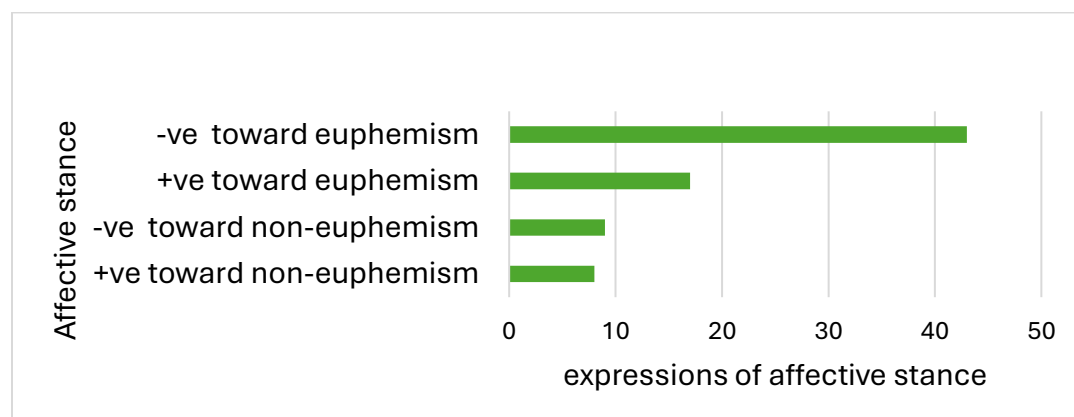
(12) Absolutrely hate it when people say "passed", "passed away" "passed over". I find it insulting.

(13) I prefer taking a dirt nap over passed away. Passed away makes no fn sense

(14) Passed what? An exam? A highway exit?

Figure 1

Affective stances toward variants of 'DIE' expressed by commenters (n=77)



Two factors may drive the dispreference for euphemism seen in the dataset. One is the association of many euphemisms with afterlife beliefs and by extension religion, with which some commenters may have had negative experiences. Another is that commenters may feel that use of euphemisms, in making denotative meaning less available or implying that things “aren’t so bad” because their loved one has just “gone somewhere else,” inappropriately deny the painful reality that those affected by death experience. This may be the case for the commenter in (12). It is also notable that dysphemisms like *take a dirt nap*, *pushing up daisies*, or *croaked*, foreground the physical reality of death, and perhaps it is this acknowledgement that makes them more acceptable than euphemisms to some.

Commenters differed in their opinion of whether use of euphemisms over non-euphemisms help those experiencing grief. Roughly two thirds of those who discussed this thought euphemisms helped grieving people, particularly in the midst of interaction as expressed by the commenter in (15). Another third considered euphemisms unhelpful, as in (16), or even detrimental, particularly to the longer-term process of working through grief (17). This is consistent with the majority and minority opinions on euphemism and grief reported by Rawlings and colleagues (2017).

- (15) It’s [...] less likely to cause people to become sad or make them remember everything all over again.
- (16) Saying “passed (away)” isn’t going to make anyone’s grief less painful.
- (17) As a bereavement counsellor I say died. Using euphemisms is unhelpful and should be avoided especially when talking to children or in when informing someone of a death. People need to know what has happened to start processing.

Some commenters expressed the belief that use of euphemism is avoidant behaviour motivated by a discomfort with or fear of death, as in (18) and (19). This after all fits within classic definitions of euphemism, such as “[...] the means by which a disagreeable, offensive or fear-instilling matter is designated with an indirect or softer term” (Kany, 1960, p. v).

Alternatively, humour, an effective strategy for managing mental-emotional discomfort (Moran, 2003; Strick et al., 2009), may motivate use of dysphemisms perceived as clever or funny as well as references to comedic entertainment media in the data set (20)².

(18) People are terrified of dying. Passed away makes it seem not that bad. Ignorant people.

(19) It's just the euphemism train. Death is uncomfortable so we'll say anything but it.

(20) Dropped off the twig 🐦

Despite this conjecture, how significant is individual fear or discomfort with death as a motivator of variant choice? The commenter in (18) is, after all, speculating on the motivations of the speakers they judge. The commenter in (21), in contrast, makes a point of stating that his preference for euphemism is not fear-based. He also implies that his perception of *passed* as 'soft' is related to his status as a native speaker of English.

(21) My point? I am not afraid of death, trust me. But 'passed' is a softer word to those who have English as their native language.

This suggests that the affective meaning of euphemisms like *PASS* is not only a consequence of their indirect nature, but also imbued by socialization in the language in question. Perhaps discomfort with death on a societal level has led to the normative use of euphemisms for 'DIE' in many English-speaking communities, but the socially constructed connotative and affective

² Example (20) refers to a Monty Python sketch where two characters argue about whether or not a parrot is dead using both direct and creatively indirect language.

meanings they carry as a result now also motivate their use, and *individual* fear or discomfort need not be involved.

2.6 Relevance of communicative context

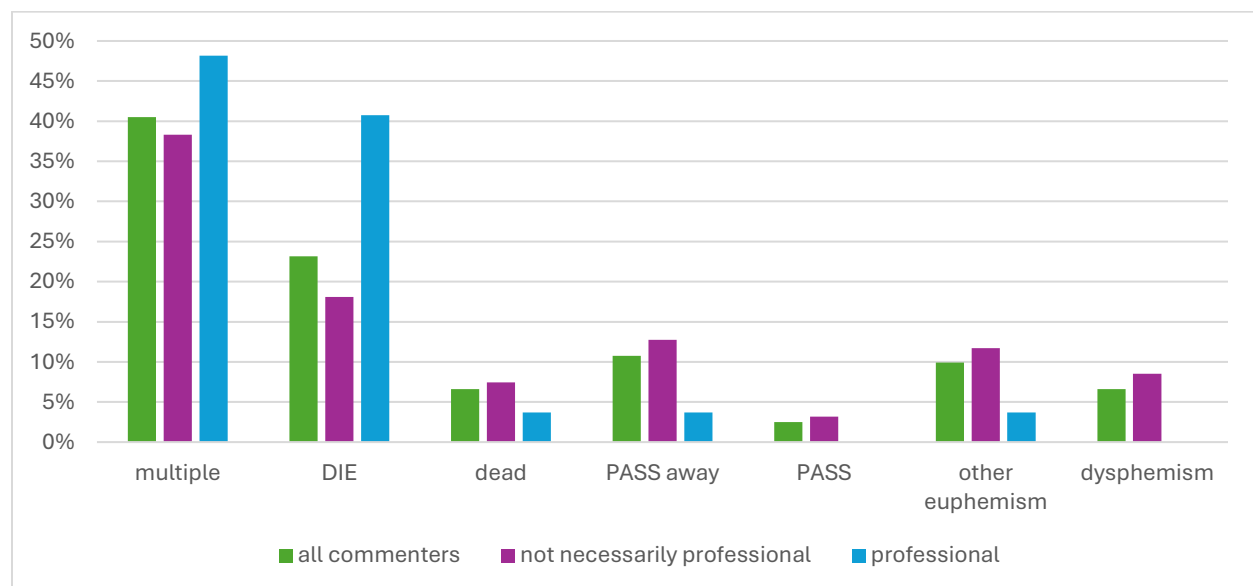
One hundred twenty-one commenters indicated what variant or variants they used, which is quantified in Figure 2. The most common behaviour reported was the use of multiple variants, particularly among commenters who were required to discuss death in a professional capacity. Commenters indicated that their use of multiple variants for 'DIE' was context-dependant (22). Further, example (23) suggests that matching the nuances of variant meaning discussed in previous sections to the situation at hand is part of what guides variant choice.

(22) I say both. It just depends on the conversation and who I'm with.

(23) I like to change it up; part of the beauty of English is the ability to use words that mean the same thing but have differing connotation for different circumstances.

Figure 2

Reported use of variants of 'DIE' by commenter (n=121)



Relevant aspects of the communicative context include whether the purpose of communication is primarily informative or supportive, the amount of social distance between the speaker, addressee and deceased and whether the setting demands formality/politeness.

In communicative contexts whose main purpose is to convey information, non-euphemisms appear to be preferred, since they provide maximum availability of denotative meaning (i.e. clarity). Such contexts include journalism, where use of *DIE* is a stylistic norm mediated by industry guidelines, as described in example (24). It also includes informing next of kin that a loved one has died. Here availability of denotative meaning is also key, as described in example (25), but also harsh affective meaning. Commenters like the one in example (26) felt that this harshness is part of what allows the information to cut through in highly stressful and emotionally overwhelming situations. Unlike reports in the literature where medical professionals most often used euphemisms when discussing the possibility of patient death with families or informing colleagues of a death in writing, most professionals in the corpus reported using non-euphemisms when informing next of kin of a death.

- (24) [...] Associated Press style (which is the default standard for journalists across most of the world) is very insistent on never using euphemisms and always saying someone died in order to provide maximum clarity [...]
- (25) Don't use vague terms. Their brains are struggling enough. Keep the communication clear. It doesn't mean that you can't show sympathy, just that you want the message to be clear
- (26) You almost have to say it kind of bluntly to really drive the point home because some people don't want to believe you and will look for any way to argue that you're wrong.

In contrast, supportive communicative contexts may not involve conveying any information that is new to the addressee. Commenters emphasized that their word choices in these situations were about what the grieving interlocutor might want rather than their own preferences (27), but that since this is generally guesswork, favouring euphemisms, with their softer affective meaning, was erring on the side of caution (28).

(27) Died. I don't find it harsh. Some people do, so I try to read the room.

(28) Death does not deserve baby words, but it does deserve tact. Because you never know what is needed in any situation, so better to go into it as carefully as you can.

Moving beyond the dimension of communicative purpose, in speech that reports a death, social distance will vary between all three (or more) individuals involved: the speaker, the addressee(s), and the deceased. Commenters generally agreed that greater social distance between the deceased and the other interactional participants, as with a celebrity for example, favours use of non-euphemism, as in (29). Commenters also felt that social distance between speaker and addressee affected their word choice, but whether smaller social distance favours euphemism or non-euphemism seemed to vary by individual as in (30) and (31).

(29) I only used died if it's people I don't know. Like random celebrities deaths.

Passed away if it's someone personal to me or someone I know.

(30) Speaking to or in front of family or close friends: passed away.

(31) But with friends and family, it's usually 'died.' No need for euphemisms there, just straight talk.

There were also commenters who reported use of different non-euphemistic variants from standard and non-standard varieties of English according to social context, as we see in (32) and

(33). For these speakers the standard English variant *DIE* was reported for professional contexts or more socially distant addressees, and *dead* with either copula deletion (a feature of both Southern American English (SAE) and African American Language (AAL); Wolfram, 1974) or invariant aspectual *be* (a feature of AAL; Green, 2000) with a close addressee.

(32) They ded-someone close to me

They died-professionally

(33) I use dead/died when people ask if parents are still alive.

“No, they died”

Or if a friend asks: nope, they be dead

Closely tied to both social distance and setting are notions of (in)formality and (im)politeness. Commenters agreed that some variants of 'DIE' are more formal than others, but, as with social distance and variant choice, fell into two opposing camps regarding which variants are more appropriate in formal contexts. One group, exemplified in examples (34) and (35), felt euphemism, especially *PASS*-type euphemisms, were more formal.

(34) I alternate between both, but I use passed away in more formal settings.

(35) I think of "passed on" as something said in casket advertising or formal settings or fancier company, maybe an Episcopalian or Catholic funeral.

Others, represented in (36) and (37), considered direct language most appropriate for formal contexts. It is possible that this dimension of meaning varies geographically. The commenter in (37) appears to describe Americans from an outside perspective and so is likely representing a non-American viewpoint. *Twee*, 'affectedly or ridiculously quaint or sentimental', employed in example (36) is a chiefly British term (Collins, n.d.; Merriam-Webster, n.d.). Taken together this

would support a hypothesis that euphemisms are formal in North America, but informal in the UK, though there is far from enough evidence to draw a definitive conclusion.

(36) Because euphemisms are twee and not appropriate for serious events.

(37) Only ridiculous Americans use euphemisms formally

There was similar disagreement as to whether euphemisms and non-euphemisms are (im)polite, expressed in comments (38) and (39) which is likely related to the differences in perception of formality described above. The commenter in (38), for example, considers *DIE* rude and associates it with extremely informal language (e.g. *what's shakin bacon*). While this may be hyperbolic, it does suggest that part of the issue for this commenter is that *DIE* is too informal, from their point of view, for the context of a funeral.

(38) stop being rude. it's just how English works. passed away is a less crude way of saying someone died. For the reverse, by your logic I should go to a funeral and say "What's shakin bacon. Your dad's dead ay mate? That sucks man. good luck" instead of "How are you doing? I'm sorry for your loss"

(39) It is not rude to say died at all. That is simply the truth.

The concept of “absolute politeness,” or the idea that any utterance can be judged as polite or impolite outside of its context has been rejected by many (Leech, 2014), nevertheless Leech makes a distinction between *pragmalinguistic politeness*, or politeness out of context, and *sociopragmatic politeness*, or politeness relative to the situation. These two conceptions are schematized in Figure 3. They are related, in that a given situation will require a certain level of pragmalinguistic politeness, and deviation from that level in either the impolite *or* polite direction will be deemed impolite from a sociopragmatic point of view.

- (41) By the way, I'm Irish and I don't think I've ever once used the expression "passed away". [next commenter] Yeah, that kind of pussyfooting is a pathology of US culture.
- (42) I've noticed a cultural difference between Brits and Americans on this terminology. For some reason in America we like to soften it by saying "passed away" as if it doesn't mean the exact same thing. In Britain, people tend to not see the point in that and just say "died"
- (43) Sadly, the land of Shakespeare has succumbed to the Americanism for death.

Commenters associated use of direct language with the upper class, and use of euphemisms with lower social classes, particularly the middle class. This may be a distinction of the past (44), but appears to resonate in the present for some (45). These commenters may also be principally British, although one American commenter indicated a class distinction was consistent with their experience.

- (44) The distinction is a bit outdated now, but historically the British upper class has frowned upon euphemisms like that. Being on the top of the social hierarchy, you didn't need to temper your speech to avoid being perceived as coarse by social superiors, and doing so was seen as characteristic of the middle class.
- (45) Personally, I look down a little on those who over-euphemise too. It's very middle-class (he says in his best Dowager Countess guise). The patricians and the plebs don't give a fuck and just say what they mean, it's everyone in the middle who frets and flops.
- (46) "Passed away" or "passed on" is decidedly Non-U. "Died" is U. Thus spake Nancy Mitford.

The concepts of “U” and “non-U”, referenced in (46) stand for “upper-class” and “non-upper class” respectively and have their origin in 1950s academia (Ross, 1954), but were popularized by the upper-class English author Nancy Mitford (1956). In this context, “non-U” referred to upwardly mobile standard English-speaking members of the middle or lower classes, who might be “mistaken for upper class” if not for subtle lexical clues in their speech. Non-U shibboleths were words used by non-U speakers to sound more “refined”, where members of the upper class would normally use plain language, for example, *serviette* rather than *napkin*, *expecting* for *pregnant* and, notably, *PASS on* for *DIE* (Buckle, 1980; Mitford et al., 1956).

This is reminiscent of “hypercorrection” in Labov’s 1966 study of /r/ pronunciation in New York, where lower-middle-class speakers r-dropped less than upper-middle-class speakers, but on a lexical rather than phonological level (2006). In both cases members of the second highest social class used variants they perceived as prestigious more frequently than the highest social class. Nevertheless, this pattern may now largely be obsolete in Britain, where a great deal of both regional and social dialect leveling has taken place since the 1950s and historical upper-class norms (of pronunciation at least) are no longer considered socially desirable nor synonymous with Standard Southern British English (Lindsey, 2019; Williams et al., 1999).

2.8 Conclusion

Within the corpus examined, commenters expressed differing attitudes toward variants of ‘died,’ with the most salient factor being whether the variant was euphemistic or not. Commenters considered variants of ‘died’ to share a denotative meaning, but vary in the accessibility of that meaning. Further, variants carried opposite connotations regarding manner of death (slow in coming versus sudden) and the existence of an afterlife (implied versus not

implied) depending on whether they were euphemisms or non-euphemisms, as well as similarly opposed affective meanings (soft versus harsh).

Euphemisms were not universally preferred. Many commenters expressed a negative affective stance towards euphemisms, presenting them as unpleasant, illogical or ridiculous. Despite this, more commenters considered euphemisms to be helpful to those who are grieving than did not. Discomfort with death on a societal level rather than a personal level was seen as a motivator of use of euphemisms, and potentially of some dysphemisms, where humour is an alternative strategy to avoidance.

Contexts where the primary purpose of communication is to inform favour use of non-euphemisms for the sake of clarity, while supportive communication favours euphemism. Non-euphemisms appear preferred when the deceased was not known personally by either the speaker or the addressee, but whether a close relationship between speaker and addressee favours euphemism or non-euphemism varied by speaker. Speaker-addressee closeness may however favour use of *dead* with SAE or AAL grammar. Whether euphemisms or non-euphemisms are more formal or pragmalinguistically polite may vary geographically, but beyond this, the sociopragmatic politeness of a variant is context dependant. Use of euphemistic variants was associated with America (in contrast to the UK and Ireland) and, at least historically, with the middle class.

The use of social media commentary as the source of this data means that commenter demographics are largely unknown. On average, Reddit users skew male (60%), American (~50%) and young (46% under 30 years old in US; Ceci, 2025a, 2025b, 2025c). Consequently, these results cannot be generalized to any whole population. While this is a limitation, previous studies from the medical literature intended to represent general populations tend to enroll older,

female participants, so perhaps this project could be considered to complement previous research from a demographic perspective.

While this project sheds light on cognitive, affective and behavioural attitudes toward variants of 'died,' behavioural attitudes reflect predisposition and intention, and may not match behaviour itself. Study of use in natural speech through data drawn from spoken corpora could address behaviour directly and further elaborate the contextual factors that condition use of individual variants of 'DIE', as well as their social meanings.

Chapter 3: “At some point there's nobody there and the individual passes”: a corpus-based study of lexical variants of ‘DIE’

3.1 Introduction

This chapter moves from attitudes and beliefs about variants of ‘DIE’ to an investigation of how they are used in natural language. I take a variationist perspective, quantifying occurrences in North American spoken English corpora as well as the linguistic and social factors that might condition their use to answer the following research questions:

RQ1: With what relative frequency do the most common variants of ‘DIE’ (*DIE*, *PASS*-type euphemisms) occur in spontaneous media discourse?

RQ2: Does use of variants in this context differ according to linguistic factors?

RQ3: Does use of variants differ according to social factors?

RQ4: Does use of variants differ between Canada and the United States?

To contextualize my inquiry, I begin by reviewing the role of corpora (3.2) and lexical variation (3.3) in sociolinguistics research, before examining previous work that sheds light on ‘DIE’ as a sociolinguistic variable (3.4). Section 3.5 details how data was extracted and coded, including the linguistic and social factors considered as potential predictors of variation, as well as statistical methods employed. Section 3.6 presents quantitative results, while section 3.7 explores their implications. Section 3.8 concludes.

3.2 Corpora in sociolinguistic research

A corpus is a large, systematic sample of language assembled as a basis for answering research questions (Egbert et al., 2022, pp. 2–11; McEnery & Brezina, 2022, p. 3; McEnery & Hardie, 2012, pp. 1–15; O’Sullivan et al., 2025, pp. 9-). Most corpora, whether made up of signed, gestured, spoken or written texts, are designed to be machine-readable in order to

facilitate analysis of large volumes of data. Although some corpora contain audiovisual files or phonetic transcription, written texts are most common, followed by spoken data rendered in standard orthography. Corpora may be annotated with linguistic information (e.g. part of speech (POS), lemma) and speaker metadata (e.g. speaker ID, sex, age, social class, region) or contain no annotation at all.

Many authors make a distinction between two main types of corpora, though they refer to them by several different names (Kendall, 2011; O'Sullivan et al., 2025, p. 10; Oushiro, 2023). Conventional, standard or general corpora collect naturally occurring language that is balanced by genre and considered representative of a whole language or variety (e.g. Corpus of Contemporary American English (COCA), British National Corpus (BNC)). They have traditionally been designed and compiled by corpus linguists (Hansen, 2018; Oushiro, 2023). In contrast, unconventional or specialized corpora capture language use in specific contexts and according to locally relevant categories, like communities of practice (e.g. *NoTa-Oslo* an audiovisual corpus of informal interparticipant conversation in Oslo Norwegian, Stratton & Sundquist, 2022). They are orders of magnitude smaller than general corpora and more likely to represent non-standard varieties because of their specialized nature. In addition to naturally occurring language, specialized corpora may contain data gathered in sociolinguistic interviews or through learner elicitation tasks and be compiled by linguists outside the field of corpus linguistics, for instance, sociolinguists or child language acquisition specialists. O'Sullivan and colleagues (2025) make a further distinction, considering both general and specialized corpora to be "conventional," in that both types are assembled for linguistic research, and use the term "non-conventional corpus" to refer to a body of texts gathered for some other purpose that can

nevertheless be used for linguistic research (e.g. Winnipeg Free Press archives, Ford & Loureiro-Rodríguez, forthcoming).

Sociolinguists often assemble their own specialized corpora to target populations or variables of interest (e.g. the corpus of Toronto Youth English, Ontario Dialects Project) (Kendall, 2011; Tagliamonte & D’Arcy, 2004; Tagliamonte & Jankowski, 2019). Nevertheless, large general corpora have been utilized for sociolinguistic inquiry, most often investigating variation according to sex, gender, age, and social class in corpora that are annotated for these speaker characteristics (e.g. Michigan Corpus of Academic Spoken English, the demographically sampled subset of the spoken BNC94, BNC14; (Aijmer, 2018; Rayson et al., 1997; Säily, 2011; Schmid, 2003). Comparison of whole corpora or subcorpora has also been used to investigate differences between language varieties, registers and time (Baker, 2010b; Hansen, 2018; H. M. Lehmann, 2002; Sigley & Holmes, 2002; Zhou & Xia, 2024).

While large, general corpora are a rich source of natural language data, their use for sociolinguistic inquiry is not without challenges. First, retrieving the social characteristics of speakers, or even unique speaker IDs may be impossible or involve time-intensive hand coding, since the majority of large corpora are not tagged with social metadata (Kendall, 2011; O’Sullivan et al., 2025, p. 30-31). Even when provided, this coding is often limited to the broadest macrosocial categories like sex and age. Apart from labour, this presents a challenge for interpretation of results. Without speaker IDs, outliers cannot be identified, nor can their effects be mitigated by including random speaker effects in multivariate models. In this situation a highly atypical speaker or unaccounted-for subgroup can bias group results (Baker, 2010a, p. 56; Brezina & Meyerhoff, 2014). For example, in BNC32³ a minority of speakers use *FUCK* (28%),

³ A socially stratified subsample of 32 speakers from the spoken BNC.

but many who do so use it copiously. Brezina and Meyerhoff (2014) show that without taking individual speaker differences into account, male gender, younger age, working class and Southern region significantly predict use of *FUCK*, yet when individual differences are accounted for only younger age remains significant.

Second, interacting with datasets via query (because they are too large to read in full) may require variationist studies to compromise on the principle of accountability, which requires that all variants of a variable be reported, including non-occurrence (Labov, 1982, p. 30). In other words, only previously identified variants can be queried, and zero forms present a unique problem. For example, in order to identify zero subject relatives (like example 1), Lehmann (2002) queried the POS string V_{finite} NP V_{finite} , which returned 20,000 results. He determined by manual inspection that only 299 of these were zero subject relatives. Even with this level of care, Lehmann conceded that he could not have identified every zero subject relative in the dataset as they also occur in other less common configurations (2).

- (1) We’ve got a cat $\emptyset$ comes to our house, a white one. (BNC:KD8, Lehmann, 2002, p. 168)

V_{finite} NP V_{finite}

- (2) To this date I still haven’t seen an Epsom print, including the 9000, $\emptyset$ looks better than my Tektronix 440 dye-sub print. (rec.photo.digital post, Lehmann, 2002, p. 169).

Finally, while general corpora are built to be balanced and representative, one must keep in mind just what they represent—standard varieties of language. Kendall points out that “in attempting to represent national varieties, [large general corpora] are by necessity somewhat normative and exclusive—they downplay and/or normalize over the true diversity of language at a national scale” (2011, p. 363). Indeed, the principle of balanced representation is typically implemented by stratification according to genre or register in general corpora, rather than by

social factors as is the norm in sociolinguistics (cf. BNC32, Brezina & Meyerhoff, 2014). In other words, there is no systematic attempt to ensure that speakers of all backgrounds are represented in general corpora, and privileged voices, including those of standard language users, are likely to be overrepresented (Hansen, 2018, p. 118). Even some spoken subcorpora, which we might expect to contain a greater variety of norms than written texts, lean toward standard language (O’Sullivan et al., 2025, p. 30,38). For example, Morin and Grieve (2024) found exceptionally few double modals in the spoken subcorpus of COCA compared to a corpus of American tweets, a difference they attribute to the relative formality of the television and radio interviews that make up COCA:spoken.

Despite these challenges, corpus data are well-suited for answering my research questions. In North America today, death is largely hidden from daily life (Aries, 1974; Arnup, 2018; Gorer, 1955; Walter, 1994). This makes death an infrequent topic outside of specialized communicative contexts like end-of-life care, and obituaries where ‘DIE’ has been studied in the past (Barlet et al., 2022; Crespo Fernández, 2006; Ferraro, 2019; Ford & Loureiro-Rodríguez, forthcoming). Because my goal is to examine lexical variation of ‘DIE’ in everyday, spontaneous spoken discourse, a deep source of searchable, naturally occurring spoken language is required to obtain sufficient tokens for quantitative analysis. This makes the spoken component of a large, general corpus an appropriate data source. While many such corpora lack speaker metadata, extracted tokens can be coded for any linguistic or social factors of interest available from context or external ratification. The fact that large, general corpora aim to represent whole national varieties of language is in this case a strength, and would facilitate comparison of how ‘DIE’ is used in Canada and the United States.

3.3 Lexical variation

Central to the study of linguistic variation and change is the concept of the sociolinguistic linguistic variable, or a set of “alternative ways of ‘saying the same thing’” conditioned by both linguistic and social (also known as internal and external) factors (Labov, 1969, p. 738). While linguistic variables can exist at any level of grammar (Wolfram, 1993, p. 195), the focus of sociolinguistic variationist scholarship has been on phonetics and phonology, and to a lesser extent morphology and syntax (Sandow et al., 2026). Comparative lack of engagement with lexis is partially due to an assumption that words do not vary systematically (Adams, 2014, p. 164; Labov, 2012, p. 16; Trudgill, 1986, p. 25), but also due to genuine methodological and theoretical challenges that are particular to lexical variation (Sandow et al., 2026).

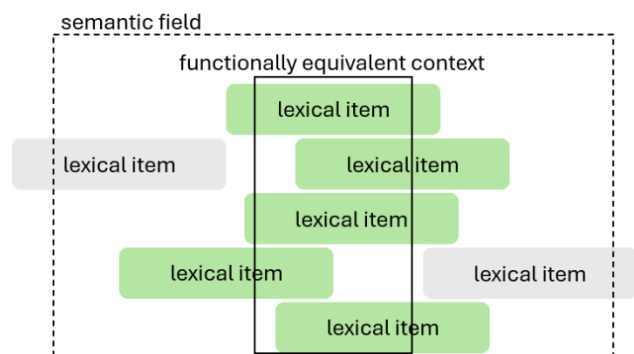
First, variants of a lexical variant may not occur with sufficient frequency for systematic analysis in sociolinguistic interviews (Labov, 1972, p. 8; Sandow et al., 2026), the standard source of data in variationist study. This is especially true of the most content-full words from open-class parts of speech. As a result, many of the lexical variables studied are those that fall toward the more functional end of the spectrum and therefore occur with higher frequency (e.g. intensifiers, Ito & Tagliamonte, 2003; quotatives, Gardner et al., 2021; pronominal quantifiers, Jankowski & Tagliamonte, 2021). To study less frequently occurring lexical variables sociolinguists have assembled or combined ever larger collections of sociolinguistic interviews (Tagliamonte & Jankowski, 2025), but also drawn data from alternative sources like large general corpora (Aijmer, 2018), computer mediated communication (Eisenstein et al., 2014; Grieve et al., 2018), and direct elicitation (Sandow, 2024; Stamp et al., 2014).

Second, lexical variation adds a layer of theoretical and methodological complexity to circumscribing the variable context and following the principle of accountability (Sandow et al.,

2026). In contrast to sets of phonological variants, which don’t in themselves carry denotational meaning, every lexical variant has a unique etymological history and range of uses. To say that two lexical items “mean the same thing,” is by necessity an approximation (Edmonds & Hirst, 2002; Goodman, 1949). Because of this, lexical variationists take an onomasiological⁴ approach, identifying members of a semantic field that express the same concept in a given context and are therefore *functionally* equivalent in that context (Dines, 1980; Stratton, 2023). Figure 4 conceptualizes this process, showing excluded members of the semantic field that are not equivalent in context in grey. Retained lexical variants (green) exhibit “weak complementarity” of distribution: although contexts may exist where one variant may appear, but not another, in the relevant context all may appear and as the rate of one variant increases the rates of others decrease⁵ (Sankoff & Thibault, 1981, p. 207).

Figure 4

Circumscribing the variable context of a lexical variable (variants in green)



Even so, the methodological challenge of identifying and collecting *all* variants that meet such criteria remains. For relatively small collections of sociolinguistic interviews, it is possible to

⁴ Onomasiology starts with a concept and looks for how it is expressed in contrast to semasiology that starts with a word and investigates its possible meanings (Stratton, 2023).

⁵ This contrasts with strong complementarity where two forms never appear in the same context.

read all transcripts and manually identify variants as they occur (e.g. quotatives in the corpus of Toronto Youth English, Tagliamonte & D’Arcy, 2004). When the size of a data source necessitates interaction by query, other strategies must be employed. One approach is to query the variable context rather than the variants themselves. This has been applied fruitfully in studies of intensifier variation that extract intensifiable adjectives from POS-tagged corpora (Brook et al., 2025; Stratton, 2022; Stratton & Sundquist, 2022). When it is not possible to query the variable context, lists are generated from dictionaries and thesauruses with the intent to exhaustively identify and query all functionally equivalent variants in a semantic field (Franco & Tagliamonte, 2021; Jankowski & Tagliamonte, 2019; Stratton, 2022, 2023; Tagliamonte & Brooke, 2014; Tagliamonte & Jankowski, 2019; Tagliamonte & Pabst, 2020; Tagliamonte & Smith, 2021). For example, Tagliamonte and Pabst (2020) used the Oxford English Dictionary Thesaurus and Urban Dictionary to identify adjectives of highly positive evaluation (e.g. *cool*, *awesome*, *brilliant*).

Finally, there is compromise and questioning. When studying lexical variation in words for treed areas, Tagliamonte and Jankowski (2025) identified 12 potential variants, but only extracted and analyzed the three most relevant in Ontario English, *woods*, *bush* and *forest*. Similarly, of 231 potential vocatives, Needle and Tagliamonte (2025) retained only the seven most frequently occurring variants to form a dataset for analysis. Taking a different tack, Aijmer (2018) constrained her study of intensifiers to those that appeared with higher frequency in BNC14 compared to BNC94. Still others have questioned the principle of accountability itself, arguing that individual variants carry social meaning regardless of whether they are contrasted with alternatives (Campbell-Kibler, 2010).

While navigating these challenges, lexical variation studies have contributed to the ongoing development of (socio)linguistic theory. They inform our understanding of delexicalization and grammaticalization (Cheshire, 2007; Johnston et al., 2015; Tagliamonte & Roberts, 2005), the constant rate effect (Gardner et al., 2021), dialect leveling and language contact (Stamp et al., 2014), the role of gender in linguistic variation and change (Aijmer, 2018; Franco & Tagliamonte, 2021; Tagliamonte & D’Arcy, 2004; Tagliamonte & Jankowski, 2019), as well as factors that affect linguistic diffusion including geographic proximity, population density, and sociocultural similarity (Eisenstein et al., 2014; Grieve et al., 2018; Jankowski & Tagliamonte, 2019, 2021; Tagliamonte & Jankowski, 2019). Others have noted the role of indexical meaning of variants in identity construction and stancetaking (Ford & Loureiro-Rodríguez, forthcoming; Sandow, 2024), though this is more common in studies that focus on a single lexical item (Denis, 2021; Kiesling, 2004). There is ample precedent that studying lexical variation is worthwhile and, given the rich and ever-changing nature of the lexicon, much remains to be discovered in this domain. Verbs that mean ‘DIE’ are one such compelling case for examining lexical variation.

3.4 ‘DIE’ as a linguistic variable

‘DIE’ differs from previously studied linguistic variables, in that its variants consist of a direct language variant *DIE* that has been in use since at least Middle English (Oxford University Press, n.d.-a) and a constellation of euphemistic variants prone to more frequent renewal due to the euphemistic cycle (Henderson Taylor, 1974). For example, participants in a Massive Open Online Course (MOOC) were able to generate 1183 unique English euphemisms and dysphemisms meaning ‘death’, ‘dead’ and ‘die’, including those that are no longer in common use (Rawlings et al., 2017). An entry point to understanding this variation is the choice between

direct language and euphemism. In contemporary obituaries, direct language is less common than euphemistic alternatives (Allan & Burrige, 1991; Cebrat, 2016; Ferraro, 2019; Ford & Loureiro-Rodríguez, forthcoming; Matiki, 2001) and the same is true when discussing death in medical practice, despite guidelines to avoid misunderstanding by using clear, non-euphemistic language (Barlet et al., 2022; Habeck-Fardy, 2021). Within this overall tendency, however, rates of direct language use vary greatly even within similar communicative contexts as seen in Table 5.

Table 5

Rates of direct language use in linguistic and medical literature

Sample	Year(s)	Rate	Reference
<i>Melbourne Sun</i> obituaries and in memoria (Australia)	1988	<1%	Allan & Burrige, 1991, p. 161
<i>New York Times</i> family death announcements (US)	2012	44%	Cebrat, 2016, p. 175
<i>Grand Forks Herald</i> obituaries (US)	Oct 2017 – May 2018	26%	Ferraro, 2019
<i>Winnipeg Free Press</i> obituaries (Canada)	2019 only	14%	Ford & Loureiro-Rodríguez, forthcoming
<i>The Nation</i> , <i>The Saturday Nation</i> , <i>Malawi News</i> obituaries (Malawi)	1996	~5%	Matiki, 2001
meetings with families of terminally ill patients (US)	2018 – 2020	8%	Barlet et al., 2022
discharge summaries (Australia)	2016 – 2019	13%	Habeck-Fardy, 2021

Factors that might condition the choice between direct and euphemistic variants have rarely been investigated. Some studies associate direct language in obituaries with unexpected deaths (Matiki, 2001; Ondimu, 2014) and male gender of the deceased (versus *passed away*, Ferraro, 2019), but these are not universally observed (cf. Ford & Loureiro-Rodríguez,

forthcoming, no effect of deceased gender or unexpectedness; Matiki, 2001, no effect of deceased gender). In the context of end-of-life care, patient’s family members use direct language more readily than clinicians (Barlet et al., 2022). Both family member and clinician use are related to clarity: family members used direct language when recounting information they had been told (presumably to confirm understanding), while clinicians used direct language in responses to family requests for clarification.

Though often grouped for analysis, euphemistic variants of ‘*DIE*’ are far from homogenous. Such euphemisms may be metonymic (e.g. took her last breath), but are more often conceptual metaphors like DEATH IS SLEEP (3), DEATH IS REWARD (4, 6), and DEATH IS DEFEAT (5). DEATH IS DEPARTURE is however the most frequent, both in English and crosslinguistically (Rawlings et al., 2017; Solheim, 2014, p. 30; Sullivan & Wachowski, 2020). Variants of this type may be creative and variable, as in (6), or have lexicalized and entered common use as with *PASS away* (Ford, 2026).

- (3) [...] peacefully *fell asleep* in the Lord (Am210a) (Solheim, 2014, p. 41)
- (4) [...] was *promoted to Glory* on Monday morning [...] (I-28) (Cebrat, 2016, p. 178; italics added)
- (5) [...] *lost her* long and courageous *battle* with lung cancer [...] (I-530) (Cebrat, 2016, p. 195; italics added)
- (6) [...] watched her *step into the presence of Jesus and into her heavenly reward*.
(COCA:NEWS 2016; italics added)

PASS away, often grouped with other *PASS*-type DEATH IS DEPARTURE variants like *PASS on*, and *PASS* alone, is the most frequently noted euphemism for ‘*DIE*’ in the literature. *PASS away* is used as a euphemistic alternative to *DIE* across the English-speaking world; its use has been

noted in English-language obituaries of postcolonial contexts like Ghana, Malawi and Kenya (Afful, 2012; Matiki, 2001; Ondimu, 2014) and it is the dominant euphemistic variant in contemporary American, Canadian, British (and possibly Australian⁶ obituaries (Allan & Burridge, 1991, pp. 161–164; Cebrat, 2016, p. 175; Ferraro, 2019; Ford & Loureiro-Rodríguez, forthcoming; Solheim, 2014, p. 31)). *PASS away* figures prominently in discussions of death language in the medical literature in the US, Australia and the UK, and when quantified, is the most frequent variant of ‘DIE’ in these contexts as well (Barlet et al., 2022; Habeck-Fardy, 2021; Rawlings et al., 2017; E. Wilson et al., 2024). Within North America, *PASS away* isn’t associated with a particular region—the Dictionary of American Regional English (DARE) describes it as widespread and standard (Cassidy & Hall, 1985-) and the Canadian Oxford Dictionary lists “*pass away* for *die*” as the sole example in its entry for *euphemism* (Barber, 2004a).

Other *PASS*-type variants like *PASS over*, *PASS on* and *PASS* alone are mentioned less often in the literature. Ford and Loureiro-Rodríguez (forthcoming) quantify *passed* separately from *passed away* in a diachronic analysis of Canadian obituaries. There it is used less frequently than *passed away*, in potentially less formal contexts. It only appears from 2007 onward, suggesting that, in Canada, *passed* is a newer variant than *passed away*. It is not surprising then that the Canadian Oxford Dictionary (Barber), published in 2004, does not list ‘died’ as a meaning for *passed* in the absence of *away*, *on*, or *over*. The Oxford English Dictionary Online (Oxford University Press, n.d.-b) does, but associates this use with North America and Spiritualism. Digging further into the American context, DARE describes *pass* as chiefly South and South Midland with frequent use among Black speakers (Cassidy & Hall, 1985). The earliest attestation listed is from a Black speaker in Louisiana in 1959. DARE also indicates that *pass on*, *pass out*,

⁶ Alan and Burridge (1991) don’t quantify variants of ‘DIE’, however *passed away* appears several times within their example sentences.

pass over and *pass off* can (or could) all mean ‘die’ in American English and, with the exception of *pass off*, are attested earlier than *pass* alone. *Pass out* is described as “somewhat old-fashioned” and, in the South, *pass on* is used primarily by white speakers. Dialectology fieldwork of the middle and south Atlantic states (see [Appendices A](#) and [B](#)) corroborates, eliciting *pass out* but not *pass* in 1930, and *pass* but not *pass out*⁷ in 1990 (Johnson, 1996, p. 272).

In sum, the literature suggests that euphemistic variants, and particularly *passed away*, are used more commonly than direct language in death-specific contexts like obituaries and end-of-life medical care. Social and pragmatic factors such as relationship to the dead or dying person, gender of the deceased, unexpectedness of death, and need for clarity may condition the choice between direct and euphemistic language. Among the most common euphemistic variants—those that contain the verb *PASS*—there is evidence that variant choice may be influenced by geographic location and ethnicity in the US and formality in Canada. No studies have yet examined how these variants are used in conversation outside of healthcare interactions, or what social and linguistic factors might condition their use in this context. My study addresses this gap by examining how ‘DIE’ is used in natural conversation in Canadian and American broadcast media to investigate factors that influence this variation.

3.5 Methodology

This project examines variation in verbs used to refer to human death in the spoken Canadian English and American English of 2019. Here I detail methodological decisions made to

⁷ The Online Etymology Dictionary suggests that the meaning of *PASS out* weakened from ‘die’ to the modern sense ‘lose consciousness’ (Harper, n.d.). In the past, *PASS out* was also used for literal motion out of a bounded space e.g. “Willard passed out, and soon returned with the ice” (COHA:FICTION; 1896), creating the basis for a DEATH IS DEPARTURE metaphor.

capture this variation, including selection of data sources (3.5.1), circumscribing the variable context through queries and exclusion criteria (3.5.2), coding of potential predictors (3.5.3) and statistical analysis (3.5.4).

3.5.1 Corpus data

Data for analysis was drawn from two corpora that represent Canadian and American national varieties of English. For American English I sampled the 127 million-word spoken subcorpus of COCA made up of unscripted conversations from television and radio programs (Davies, 2008). This large general corpus is available through the corpus tools of English-corpora.org and extends to 2019. I chose not to draw my Canadian data from its Canadian counterpart on that site, the Strathy Corpus of Canadian English (Strathy), for two reasons. First, Strathy’s spoken component is far smaller than COCA’s (5 million words) and so unlikely to yield a similar (or sufficient) number of tokens for analysis. Second, Strathy’s spoken subcorpus is not comparable to COCA’s since it contains *scripted* television transcripts, as well as parliamentary discourse, public inquiries and university lectures. Instead, I turned to the Canadian Business and Current Affairs database (CBCA), a “non-conventional corpus”, in that it is assembled as a record of Canadian business, culture, current events, education, literature, language, agriculture, history, government and politics rather than for linguistic research (O’Sullivan et al., 2025; ProQuest, n.d.). CBCA comprises millions of texts, though no exact number or word count estimate is available given that it is updated continually). These include searchable transcripts of unscripted Canadian radio and television of a similar nature to the texts in the COCA spoken subcorpus, making it an appropriate Canadian data source for comparison.

3.5.2 Circumscribing the variable context

In the context of a corpus study where data is collected by query, it is not possible to exhaustively collect all variants of ‘DIE,’ which include conventionalized variants like *PASS away* but also less-conventionalized creative variants like 7 (reprinted from 6).

(7) “[...] watched her *step into the presence of Jesus and into her heavenly reward*.”

(COCA:NEWS 2016; italics added)

Previous research suggests that, apart from *DIE* itself, *PASS*-type euphemisms, particularly *PASS away*, are the most common variants of ‘DIE’ (Barlet et al., 2022; Cebrat, 2016; Ferraro, 2019; Ford & Loureiro-Rodríguez, forthcoming; Habeck-Fardy, 2021; Solheim, 2014). In addition, the variant *PASS* without verbal particles *away*, *on*, or *over* is of particular interest, as it appears to be an incoming variant in Canadian English (Ford & Loureiro-Rodríguez, forthcoming).

Consequently, this study focuses on *DIE* and *PASS*-type variants including *PASS away*, *PASS on*, *PASS over*, and *PASS*, as variables of greatest interest.

3.5.2.1 Corpus queries

Queries detailed in Table 6 identified all potential tokens of *DIE* and *PASS*-type euphemisms in spoken texts from 2019 (the most recent year available in COCA). COCA allows search by lemma and POS; for example, “DIE_v” returned the verb forms *died*, *die*, *dies* and *dying*, but not nouns (e.g. *tool and **die*** or *advice for the **dying***). CBCA required individual queries for each form of a lemma, apart from “die” and “pass” which returned forms with and without *-(e)s*, and did not allow POS to be specified. The program Power and Politics was excluded from the CBCA search as its transcripts lacked consistent speaker attribution.

Table 6*Corpus data collection queries*

	COCA	CBCA
queries	DIE_v PASS_v	died, die(s), dying passed, pass(es), passing
time parameter	2019	1 Jan 2019 – 31 Dec 2019
other parameters	spoken subcorpus	other sources, transcript English NOT Power and Politics
search hits before exclusions	1511 concordance lines	2503 transcripts

3.5.2.2 Exclusion criteria

Potential tokens functioning as other POS (based on distributional criteria) were excluded, as were those that did not relate human death (e.g. plant and animal death, metaphorical death, *PASS* as a verb of literal motion). Tokens used in contexts likely to involve reading or reciting written materials aloud were also excluded (e.g. newsreading, political speeches). Speakers of other varieties of English (e.g. New Zealand English) were excluded, as were those whose variety of English was ambiguous or unknown (e.g. Canadian living in the United States, interviewee in Albania). Speech aired in 2019 that was known to have been recorded in an earlier year was also excluded. In the case where the same interview was aired on multiple programs, only the first occurrence in the data was collected. Details of exclusion criteria and their application are given in Table 7. After application of exclusion criteria and exclusion of three additional tokens during coding (see section 3.5.3.2) 1000 tokens remained, 349 collected from COCA, and 651 from CBCA. The dataset is more balanced by country than these figures would suggest, since there were a larger number of CBCA tokens contributed by Americans (n=126) than COCA tokens contributed by Canadians (n=2). Final totals of by country are 527 Canadian tokens and 473 American tokens.

Table 7 *Inclusion/exclusion criteria and their application*

Inclusion principle	Exclusions	Exclusion example/explanation
verbs that mean 'DIE'		
	past participle that: = 'dead' → Adj	"Yes, you know nieces, nephews, depends on how long they've been passed." "He was already passed away and he was like a doll."
	present participle that: pre-modifies a N → Adj pre-modified by Det or Adj → N	"Here is a dying man, a dead man, a Holocaust survivor [...]" "It's not the dying. It's the leaving." "[...] to commemorate his passing."
	other meanings of <i>PASS</i> ±PREP	"[...] that bill actually passed." "It might actually pass on all the passion [...]" "[...] people who were passing by [...]"
	<i>DIE</i> off/out	not equivalent to 'DIE' in that a population can die out/off, but an individual cannot
Inclusion principle	Exclusions	Exclusion example/explanation
human death	death of non-human	"And the foal was in a position trying to look at her too when they died." CBCA
	death of parts of humans	"I would later find that his brain was slowly dying."
	metaphorical death	"innovate or die" "[...] it felt like I had almost died as a musician." "I've lived and died with the team"
	hyperbole	"[...] you're going to fail and starve and die in that order if you're a freelancer." "And I would have died for that many boys in any of my classes [...]" "[...] I will stress that from now till the day I die [...]"
	proverbs and idioms	"Those who live by the sword die by the sword" "die-hard" "die on the vine"

Inclusion principle	Exclusions	Exclusion example/explanation
spontaneous spoken English	non-spontaneous reportage	news-reading expository introduction narrative voiceover translation voiceover
	prepared statement	law enforcement public statement retirement farewell speech statement delivered in court
	the words of others	letters/email/social media posts read on air titles quotations direct reported speech (DRS) ⁸
Inclusion principle	Exclusions	Exclusion example/explanation
Canadian English or American English of 2019	other varieties of English	British English Albanian interviewee in Albania, variety of English unknown
	ambiguity between Canadian and American English	Alex Trebek, who grew up in Canada, but lived and worked in the United States for many years while maintaining dual citizenship
	re-aired program	“I only found out about the suit basically two years after he died. [...] That documentary Becoming Santa was produced by The Current's Kristin Nelson with help from The Current's documentary editor, Liz Hoath. Since we first broadcast that documentary [...]” (first aired in 2017)
	archival audio within 2019 program	“I’m not going to die here. [...] That was Christine Lang speaking to As It Happens in 2004 two days after the Boxing Day tsunami.”

⁸ DRS is a selective depiction of a previous utterance that may highlight or ignore any aspect of the original (e.g. prosody, voice quality, code, gesture) including lexical choices (Clark & Gerrig, 1990; Holt, 1996). While this means that DRS “recreates” rather than accurately reproduces a referent’s speech, there is no way of knowing whether lexical choices reflect the stance and context of the current or original speaker (or both). DRS was excluded due to this ambiguity.

3.5.3 Coding

3.5.3.1 *Linguistic factors*

To examine whether the linguistic environment might influence choice of variant, tokens were coded for the following potential linguistic predictors:

1. CLAUSE: does the variant appear in a main or embedded clause?
2. MODAL PRESENT: does a modal modify the variant? (two coding schemes are tested, one that groups the semi-modals *BE going to* and *HAVE to* with true modals and one that does not).
3. MODAL: what modal (if any) modifies the variant?
4. PERFECT: does the variant appear with perfect aspect?
5. PRESENT PERFECT: does the variant appear in the present perfect, which often highlights the relevance of past events to the present?
6. PAST PERFECT: does the variant appear in the past perfect, which often situates an event within a past narrative?
7. PROGRESSIVE: does the variant appear in the progressive?

3.5.3.2 *Social factors*

Potential social predictor variables were also coded:

1. DEATH TYPE: what is the nature of the death in question? (detailed in (8))

(8)

Code	Examples
human-mediated	crime, war, suicide, overdose, withholding the essentials of life
accident/disaster	plane crash, animal attack, nuclear disaster, hurricane, sports accident
acute health event	seizure, heart attack, sudden "health complications"
chronic health event	cancer, ALS, AIDS, "natural causes", "old age"
unknown	nature of death unknown
cannot be specified	deaths of people who died in multiple ways, hypothetical or future death that may be due to any cause

2. DEATH HAS OCCURRED: is the speaker reporting a death that has occurred? In addition to "Y" (9) and "N" (10), a third code, indeterminate, captured difficult to classify situations. These include cases where death was contextually present, but whose occurrence was not expressed by the token itself (11), "deaths (of a certain nature) happen in general" statements (12), and when there was ambiguity whether a group of people were collectively near the end of life, or deaths were occurring on an ongoing basis (13).
 - (9) I lost one friend who died as a result of Alzheimer. (COCA:SPOKEN 2019)
 - (10) So he could have actually died from the phage preparation. (CBCA 2019)
 - (11) We couldn't believe anybody would get hurt or much less die. (recounting the experience of a mass shooting; CBCA 2019)
 - (12) People with disabilities die without health care. (COCA:SPOKEN 2019)
 - (13) What we're doing here, it's so urgent because people my age, we're dying. (CBCA 2019)
3. EVENT TIME (past, non-past): Does the event described take place in the past, regardless of grammatical tense? Historical present and present perfect were coded as "past" as both of these uses of the present tense relate past events.

4. DEATH RECENCY: how many years ago did the death in question occur? Where death recency was a range of years, the most recent year was entered. Where non-numerical quantities of years were indicated the conversion to a numerical equivalent in (14) was utilized. Where year of a death that had occurred was unknown death recency was designated “unknown”, while hypothetical, future or fictional deaths received a code of “cannot be specified.”

(14)

Non-numerical quantity	Numerical equivalent
a couple	2
a few	3
several	5

5. RELATIONSHIP OF DECEASED TO SPEAKER: who is the deceased to the speaker? Are they a family member, a friend, an acquaintance, the speaker themselves or are none of those relationships known to apply (other)? When the deceased had multiple roles in the speaker’s life the closest relationship was coded (e.g. colleague and friend → friend). Similarly, if multiple people with different relationships to the speaker died, the closest relationship was coded (e.g. daughter and daughter’s friend → family)
6. RELATIONSHIP OF DECEASED TO ADDRESSEE: who is the deceased to the addressee? This was coded according to the same categories used for relationship of the deceased to the speaker.
7. SPEAKER GENDER: according to contextual cues like gendered pronouns, honorifics, and names, what appears to be the speaker’s gender? This was originally coded as female, male, non-binary, and unknown. As there was only one speaker known to be non-binary

and one of unknown gender, the three tokens they collectively contributed were excluded as outliers.

8. **SPEAKER STATUS AS JOURNALIST:** was the speaker a news professional, whether the host, a reporter, or a guest who also happened to be a journalist? It is standard practice in journalism to avoid euphemisms, particularly those for death and dying (Froke et al., 2024, p. 60). For example, the Associated Press advises journalists to avoid *passed on* or *passed away* outside of a direct quotation. For this reason, journalists are expected to use *DIE* more frequently than non-journalists.
9. **SPEAKER COUNTRY:** does the speaker speak Canadian English or American English?

3.5.4 Statistical analysis

All statistical analyses were performed in R (R Core Team, 2023). These include descriptive statistics and data visualization, random forest analysis, conditional inference trees, and logistic regression analysis. For all functions that depend on random number generation `set.seed(333)` was used for reproducibility.

A random forest model for the choice between *DIE* and *PASS*-type variants as a group was constructed with the `cforest` function of the `partykit` package (v. 1.2.24, Hothorn & Zeileis, 2015). This package was selected over the alternative, `randomForest`, for several reasons. Both packages use an algorithm that randomly selects five potential predictors⁹ and grows a tree through binary recursive partitioning (then repeats), however they use different mathematical methods to do so. The `randomForest` function judges goodness of potential splits based on the purity of the resulting branches or nodes, which biases its trees toward use of predictors that have

⁹ This number can be set manually, but the default is the root of the number of predictors. In this case: $\sqrt{22} = 4.7 \cong 5$.

more levels (Strobl et al., 2007). `cforest` uses statistical tests to test the hypothesis that predictor variables and outcome variable are independent and, if this is rejected, identifies the predictor most strongly correlated to the outcome variable. Once identified, a separate statistical test splits the data according to this predictor. The function then repeats this on the data in each of the branches created. This process is not biased toward many-leveled predictors and is considered more robust (Greenwell, 2022, pp. 111–144). In addition, `cforest` can handle NA values in numerical predictors (e.g. unknown age of deceased), while the `randomForest` function cannot.

Conducting random forest analysis in addition to regression analysis conferred several advantages. First, it allowed the calculation and comparison of variable importance¹⁰ of all potential predictors, which is particularly valuable when a large number of predictor variables are considered relative to the number of observations (Tagliamonte & Baayen, 2012). Second, it allowed comparison of predictors that are highly correlated (or are even different measures of the same phenomenon), which would preclude their inclusion in the same regression model. In fact, this property of random forest analysis was used to guide subsequent regression analysis. For example, only the measure of deceased age with the greatest predictive power (as measured by conditional variable importance) was used in any regression model. Finally, random forests can handle non-linearity and complex interaction effects, identifying the importance of any variables which would be poor predictors in regression analyses for this reason.

In addition to random forest analysis, individual conditional inference trees (CITs) were also grown and visualized using `partykit`’s `ctree` function (Hothorn et al., 2006). These trees are

¹⁰ I used conditional permutation variable importance rather than marginal permutation importance, because the later inflates the importance of correlated predictors (Strobl et al., 2008, 2009).

grown using the same statistical-test-based recursive partitioning process as individual trees in the random forest, however instead of evaluating five randomly chosen variables at a time, the algorithm uses all potential predictor variables specified. While individual decision tree methods are typically 10% less accurate for prediction than ensemble methods like random forest, they are more easily interpretable (Greenwell, 2022, p. 202) and “remain one of the most flexible and practical tools in the datascience toolbox, whether for description or prediction” (p. 36).

Recognizing their limitations, I set a conservative threshold of significance ($\alpha=0.01$) and used them in a limited fashion for description, as well as to help bin many-leveled predictors.

Many speakers within the dataset contributed only one token, creating a distribution that could inflate coefficient estimates in mixed-effects logistic regression. Instead, to maintain the independence of individual observations I limited data for analysis to one random observation per speaker using the `slice_sample` function and built non-mixed-effects logistic regression models with the `glm` function. Models were constructed for the choice between *DIE* and *PASS*-type variants as a group. A single best model was selected considering the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), two standard criteria used to compare statistical models. Lower values indicate a better-fitting and simpler model. BIC penalizes additional predictors more strongly, so it was generally given precedence when the two criteria favored different models; however, AIC was also considered when BIC values were very similar. Several models considered are discussed in section 3.6.4 for transparency. Originally, I intended to construct separate regressions predicting each individual *PASS* variant separately as well, however this was not possible due to the low number of *PASS*-type variants collected.

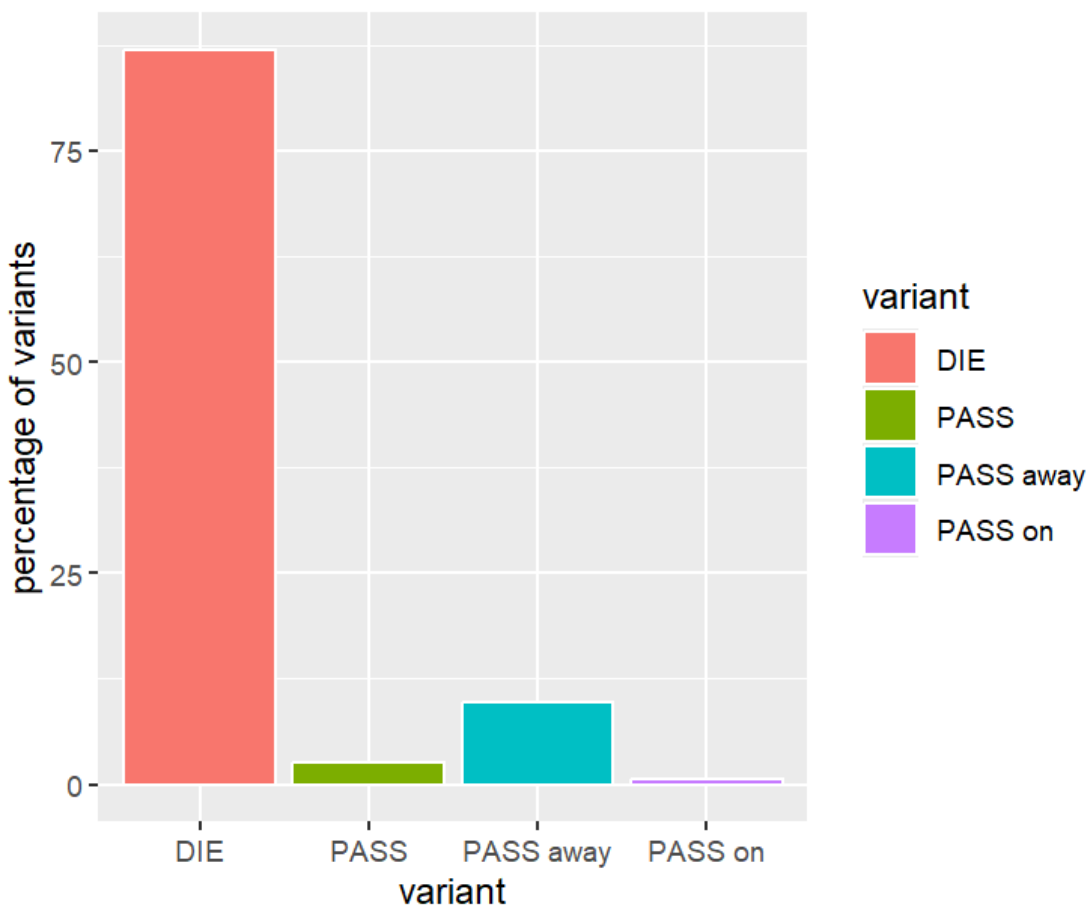
3.6 Results

3.6.1 Descriptive statistics

Overall, variants of 'DIE' were distributed in the dataset ($N = 1000$) as follows: *DIE* (87%, $n = 870$), *PASS away* (9.8%, $n = 98$), *PASS* (2.6%, $n = 26$), *PASS on* (0.6%, $n = 6$). Figure 5 illustrates their distribution. *DIE* is the dominant variant by a large margin and, as a result, the distribution is heavily skewed toward direct language (87%, $n=870$), including only 130 tokens (13%) that are euphemistic *PASS*-type variants. Among *PASS*-type variants ($N=130$), *PASS away* was most common (75.4%, $n = 98$), followed by *PASS* (20%, $n = 26$) and *PASS on* (4.6%, $n=6$).

Figure 5

Overall distribution of variants of 'DIE' ($N=1000$)



Distribution of variants by country was remarkably similar, as illustrated in Figure 6. *DIE* was used slightly more often in the US than Canada (88.2%, n=417 versus 86%, n=453). Conversely, both *PASS away* (9.5%, n=45 versus 10.1%, n=53) and *PASS* (2.1%, n=10 versus 3.0%, n=16) were marginally more frequent in Canadian speech. *PASS on*, the least common variant, occurred five times (0.9%) in the Canadian data, but only once (0.2%) in the American data.

Figure 6

Distribution of variants of ‘DIE’ by country (N=1000)



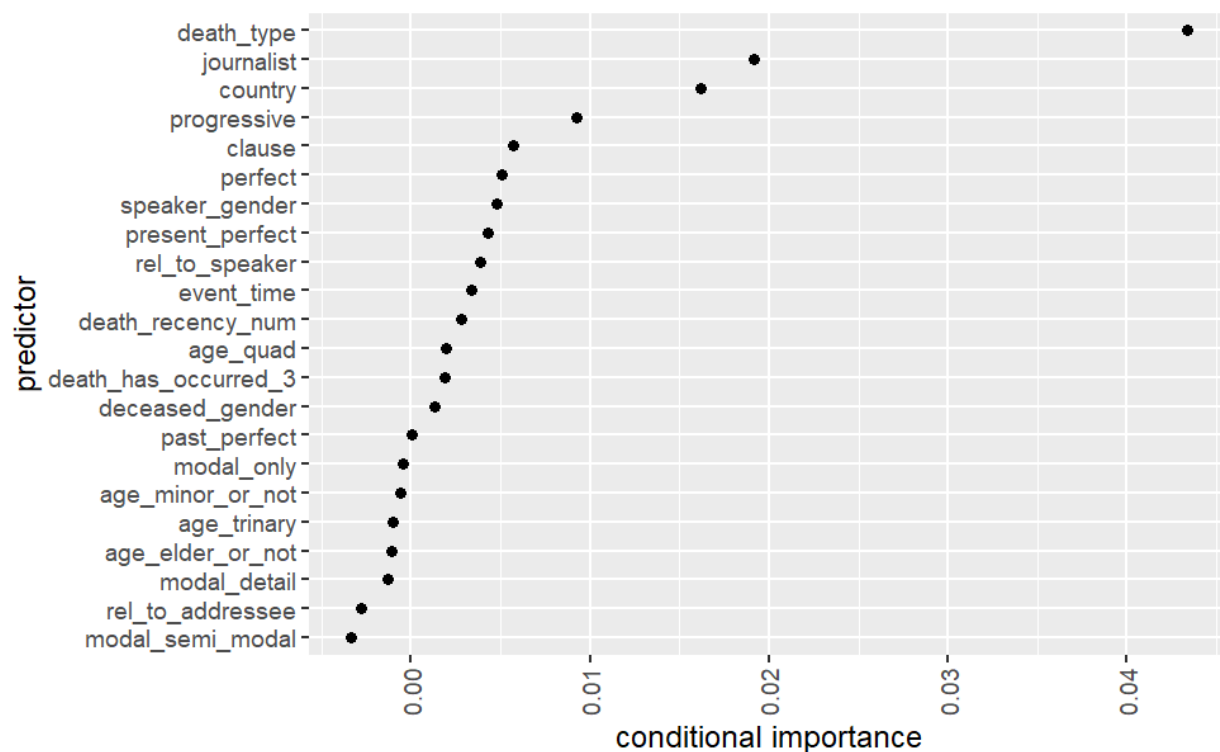
3.6.2 Random forest analysis

The 1000-tree random forest built to predict the choice between *DIE* and *PASS*-type variants was used to calculate the conditional permutation variable importance of all potential predictors. The results, given in Figure 7, indicate that DEATH TYPE is by far the most important

variable for improving accuracy of random forest-based predictions, followed by status as JOURNALIST, COUNTRY and PROGRESSIVE aspect. This analysis also tested several highly correlated potential predictors, or different ways of measuring the same phenomenon, head-to-head. Here we see that the PRESENT PERFECT and PAST PERFECT have very different variable importance (positive value, ranked 8th versus roughly zero, ranked 15th), suggesting that they should be considered separate predictors, rather than combined under PERFECT aspect. AGE_QUAD, which separates age of the deceased into the largest number of bins was more predictive than versions with fewer bins, and all attempts at measuring the effect of modals were less predictive than random noise.

Figure 7

Conditional permutation importance for the choice between DIE and PASS-type variants, all potential predictors

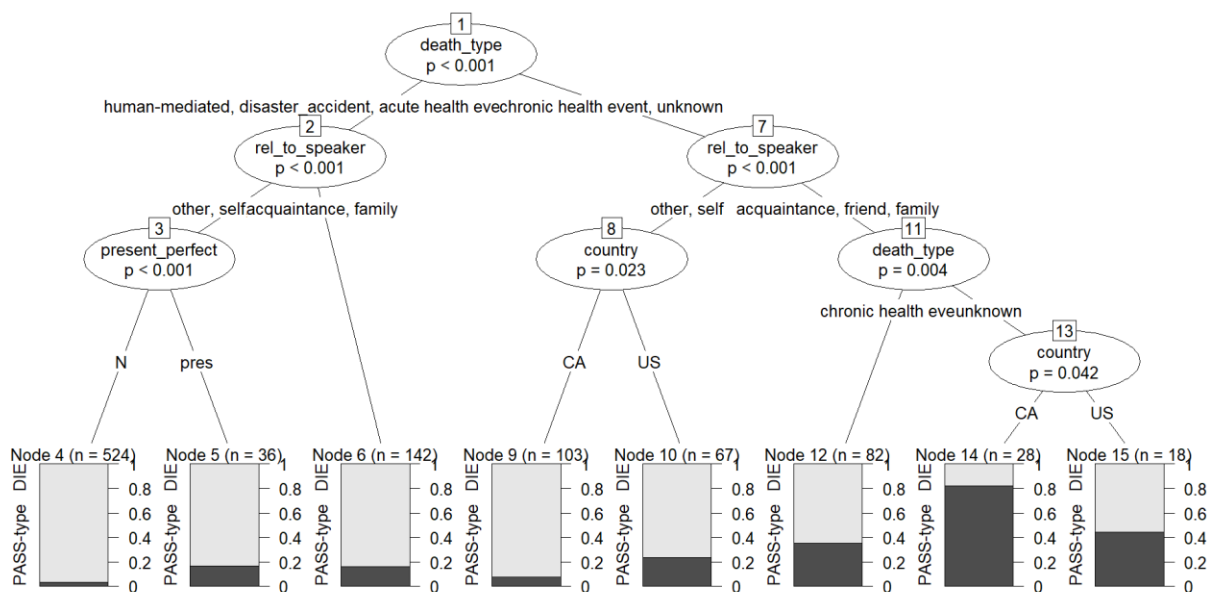


A random forest gains its predictive power by constructing many decision trees using five variables at a time and evaluating their prediction accuracy in aggregate. Figures 8a and 8b are just two examples of such trees (in this case built with the `cTree` function) and illustrate the complexity and variability of how they are fit to the data.

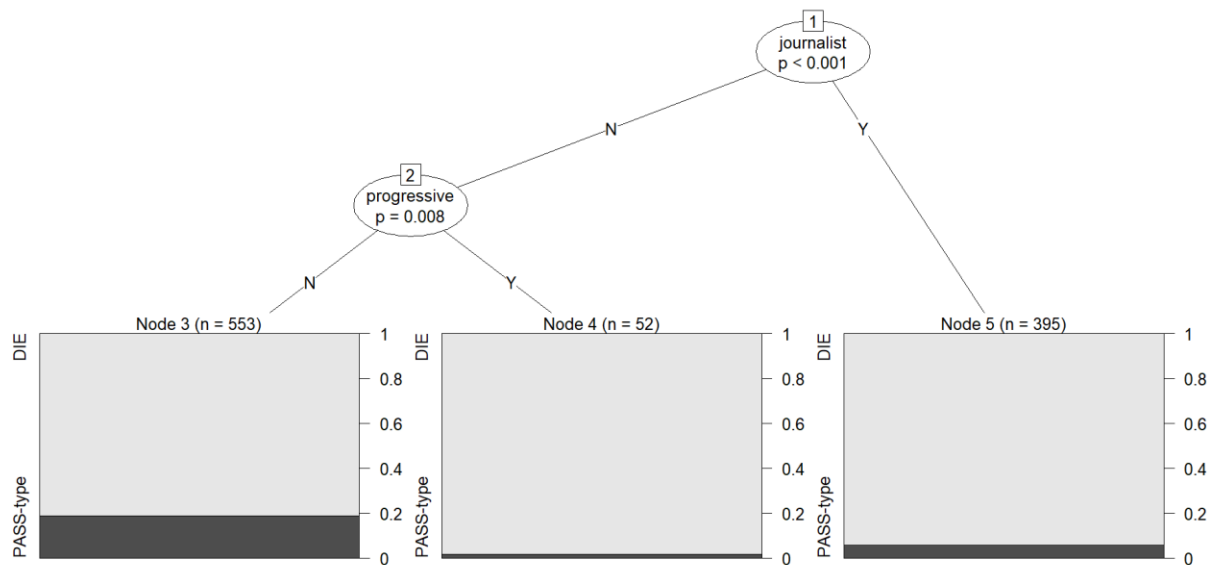
Figure 8

Example conditional inference trees from the random forest modeling the choice between DIE and PASS-type variants

a. Input: DEATH TYPE, COUNTRY, PROGRESSIVE, PRESENT PERFECT, RELATION TO SPEAKER



b. *Input: JOURNALIST, COUNTRY, PROGRESSIVE, CLAUSE, SPEAKER GENDER*



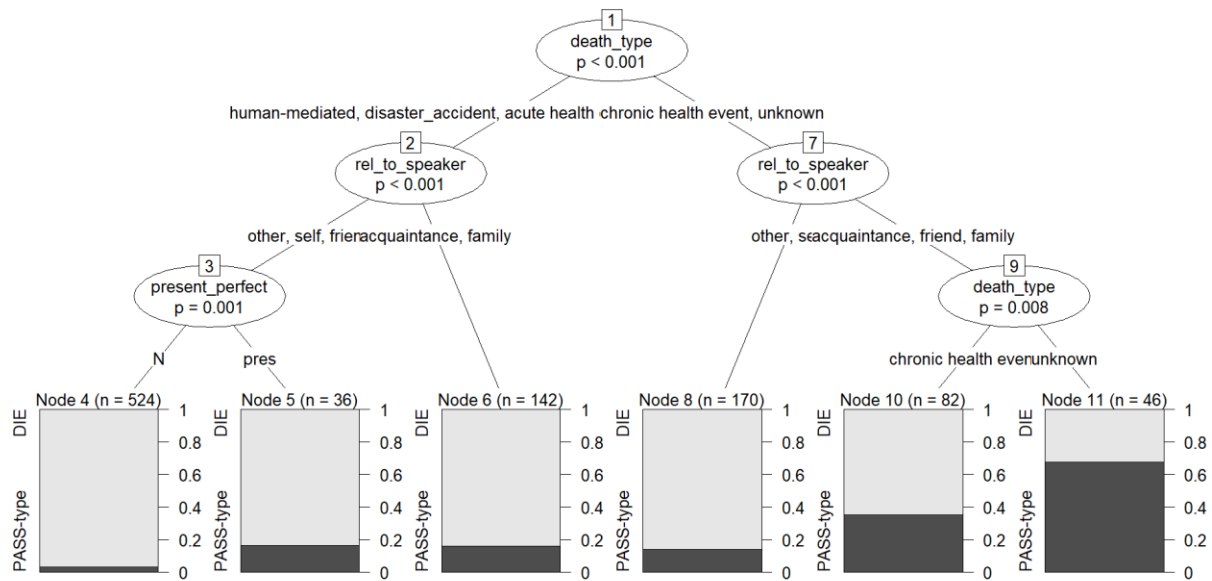
3.6.3 Conditional inference tree analysis

Unlike the 5-variable example trees in Figures 8a and 8b, the following conditional inference trees were built considering the associations of all (non-duplicate) potential predictor variables with the outcome variable. In Figure 9 we see that the predictor with the strongest association with the outcome variable is DEATH TYPE. This is no surprise, since it also had the highest variable importance. After the data is split by DEATH TYPE, however, the predictor with the strongest association with the outcome variable in each branch is RELATION TO SPEAKER, rather than predictors with higher variable importance like JOURNALIST or COUNTRY. In the right branch, which comprises unremarkable deaths, proportion of *PASS*-type variants increases across the following contexts: the death of oneself or a stranger < the death of a person known to the speaker due to a chronic health event < the death of a person known to the speaker due to an unknown cause. For sudden or violent deaths in the left branch, use of euphemistic variants was lowest for oneself, strangers and friends when not highlighted by the present perfect. When the

present perfect was used, rates were similar to those for the sudden or violent deaths of family and acquaintances.

Figure 9

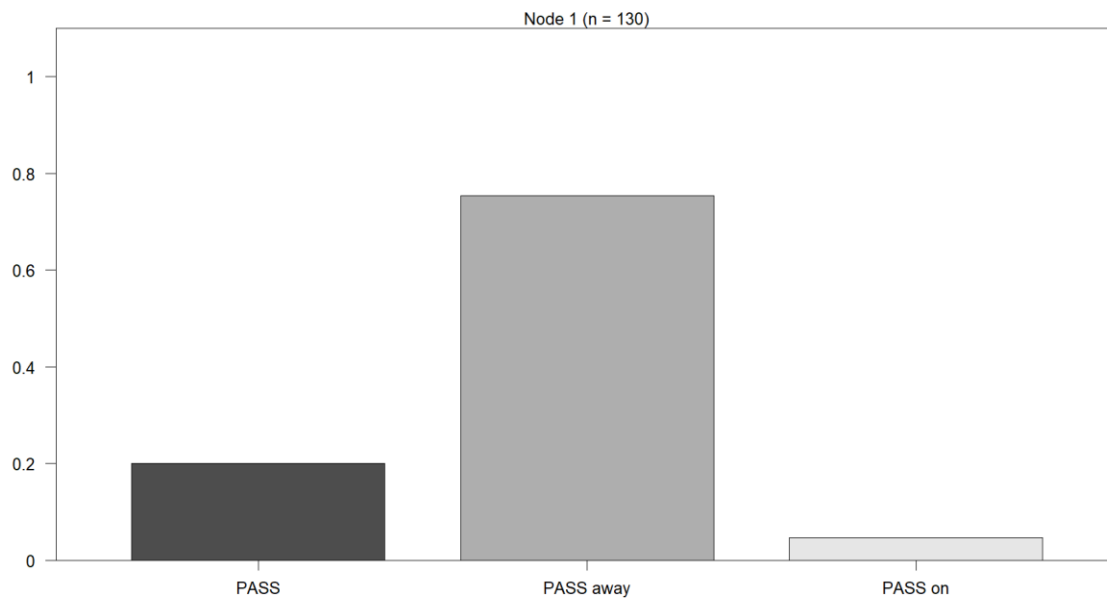
Conditional inference tree for the choice between DIE and PASS-type variants



In Figure 10, the *ctree* algorithm accepts the null hypothesis that the choice of individual *PASS*-type variant is independent of potential predictor variables and thus draws a single unsplit node.

Figure 10

Conditional inference tree for the choice between PASS away, PASS on and PASS



3.6.4 Logistic regression

For the following regression analyses I binned the many-leveled predictors DEATH TYPE and RELATION TO SPEAKER into fewer levels in order to create simpler models. Guided by the results of CIT analysis, I grouped deaths caused by humans, disaster, accident and acute health events as sudden or violent death. Deaths due to an unknown cause or chronic health condition were grouped as “unremarkable”, while “cannot be specified” remained a distinct level. Similarly, RELATION TO SPEAKER was converted to PERSONAL RELATIONSHIP and grouped as follows: “yes” if the deceased was an acquaintance, friend or family member of the speaker and “no” if they were not. It was also necessary to bin the otherwise continuous predictor DEATH RECENCY, due to unknown values and deaths that had not occurred. CIT did not identify any useful splits to guide the process, so I applied a logarithmic scale to numerical values: <1 year, 1-10 years, >10 years, unknown, cannot be specified.

The best-fit regression model (as indicated by lowest AIC and BIC) contains DEATH TYPE, DEATH HAS OCCURRED, PERSONAL RELATIONSHIP and the PROGRESSIVE as predictors. Table 8 compares the best-fit model (model 1) to models with one fewer predictor (models 2-5), one additional predictor (models 5-7), and one alternative predictor (EVENT TIME rather than DEATH HAS OCCURRED, model 8). We see that in all but one case, removing, adding or replacing a predictor does not result in any further improvement in BIC. Model 1 has a slightly higher BIC than model 4, but a lower AIC by a larger margin, and is considered the better overall model taking both measures into account. For these reasons, model 1 was selected for further analysis and is reported in greater detail in Table 9.

Table 8

Best-fit model predicting DIE (versus PASS-type variants, N=528) compared to models with one fewer, one more and one alternative predictor

Model no.	DEATH TYPE	PROG.	DEATH HAS OCCURRED	PERSONAL RELATIONSHIP	JOURNALIST	COUNTRY	AIC	BIC
1	✓	✓	✓	✓			377.8	407.6
2	✓	✓	✓				394.2	419.8
3	✓	✓		✓			395.1	416.4
4	✓		✓	✓			381.2	406.8
5		✓	✓	✓			410.7	432.0
6	✓	✓	✓	✓	✓		377.7	411.8
7	✓	✓	✓	✓		✓	378.8	413.0
8	✓	✓	EVENT TIME	✓			389.1	414.7

Compared to a baseline of sudden or violent death, *PASS*-type variants are more likely when deaths are “unremarkable” (*Est.* = -1.474, *SE* = 0.290, *z* = -5.091, *p* = 3.55×10^{-7}) or cannot be specified (*Est.* = -2.112, *SE* = 0.529, *z* = -3.994, *p* = 6.51×10^{-5}) as well as when the speaker has

a personal relationship with the deceased ($Est. = -1.202$, $SE = 0.285$, $z = -4.212$, $p = 2.53 \times 10^{-5}$). *DIE* is more likely when death has not occurred ($Est. = 1.819$, $SE = 0.520$, $z = 3.498$, $p = 2.53 \times 10^{-4}$) or falls into the “indeterminate” category ($Est. = 2.479$, $SE = 1.057$, $z = 2.346$, $p = 0.019$). Use of the progressive may also favour *DIE* ($Est. = 1.920$, $SE = 1.057$, $z = 1.816$, $p = 0.069$), as adding this predictor improved model fit and its relationship to the outcome variable approaches significance.

Table 9

Overall best-fit model predicting DIE (versus PASS-type variants, N=528)

	Estimate	SE	z	p-value estimated by glm
Intercept	2.486	0.273	9.113	$< 2 \times 10^{-16}$ ***
DEATH TYPE sudden/violent (baseline)				
unremarkable	-1.474	0.290	-5.091	3.55×10^{-7} ***
cannot be specified	-2.112	0.529	-3.994	6.51×10^{-5} ***
DEATH HAS OCCURRED yes (baseline)				
no	1.819	0.520	3.498	4.70×10^{-4} **
indeterminate	2.479	1.057	2.346	0.019*
PERSONAL RELATIONSHIP no (baseline)				
yes	-1.202	0.285	-4.212	2.53×10^{-5} ***
PROGRESSIVE no (baseline)				
yes	1.920	1.057	1.816	0.069

* $p < .05$; ** $p < .01$; *** $p < .001$

The results above provide evidence that DEATH TYPE, DEATH HAS OCCURRED, PERSONAL RELATIONSHIP and potentially use of the PROGRESSIVE matter when speakers choose a variant of ‘DIE’. When all else is held constant, *PASS*-type variants are 3 times more likely when speakers

have a PERSONAL RELATIONSHIP with the deceased, 4 times more likely when death is “unremarkable” and 8 times more likely when DEATH TYPE cannot be specified. Use of *DIE* rather than *PASS*-type variants is 6 times more likely when death has not occurred, 12 times more likely when death falls under the “indeterminate” category and, if the trend toward significance is taken at face value, 7 times more likely in the PROGRESSIVE. Below, I revisit each of my research questions in light of these results.

3.7 Discussion

3.7.1 With what relative frequency do variants of ‘DIE’ occur in spontaneous media discourse?

Overall, the use of *DIE* in the spontaneous media discourse sampled accounts for a much higher proportion of tokens (87%) than originally expected. This likely stems from the nature of the communicative context itself: programs that are focused on information, are less formal than obituaries and relate newsworthy deaths. While speakers in the dataset are generally in dialogue with one another, because these conversations are produced for an audience, their main purpose is to convey information rather than provide interpersonal support. Frequent use of *DIE* in this setting is consistent with results from my qualitative analysis (Chapter 2, section 2.6) where Reddit users report that informative communication favours direct language and supportive communication favours euphemism. Formality may also play a role. Qualitative analysis (Chapter 2, section 2.6) also suggests that *PASS*-type variants are considered more formal than *DIE* by North Americans. Within spoken contexts, broadcast media discourse is relatively formal, though still less so than the highly conventionalized genre of obituaries. This difference may help explain why obituaries of the same period use far less direct language (<1-44%; see Table 5). Finally, the nature of the deaths discussed on air skew toward what is newsworthy. This means

that a high proportion of tokens in the dataset (59%) report deaths that are sudden, violent or otherwise “remarkable”—death types that significantly favour use of *DIE* according to both CIT analysis (Figure 9) and regression analysis (Table 9).

Looking at the subset of *PASS*-type variants (n=130), *PASS* contributes 20% of tokens and *PASS away* 75%. Relative rates of these variants can be compared to 2019 obituary data from Ford and Loureiro-Rodríguez (forthcoming). In their study, *passed away* accounted for 95% (n=39) of *PASS*-type variants, and *passed* represented just 4.9% (n=2), far lower than in the media discourse analysed here. Ford and Loureiro-Rodríguez suggest that *PASS* is less formal than *PASS away*, which may account for its higher relative use in speech compared to formal obituary writing. To my knowledge, no recent studies have quantified *PASS on*, so comparison is not possible.

3.7.2 Does use of variants differ according to social factors?

The data suggests that several social factors influence variant choice including DEATH TYPE, PERSONAL RELATIONSHIP and DEATH HAS OCCURRED.

DEATH TYPE stands out as the predictor with the highest variable importance in random forest analysis (Figure 7), the most strongly correlated predictor (first split) in CIT analysis (Figure 9) and a significant predictor in logistic regression. More specifically, sudden deaths caused by human violence, disaster, accident and acute health events predict use of *DIE* rather than *PASS*-type variants. Conversely, death due to chronic health conditions and unnamed causes predict use of *PASS*-type variants. This is consistent with use of direct language for unexpected deaths in obituaries (Matiki, 2001; Ondimu, 2014) and the qualitative results (Chapter 2, section 2.4) where Reddit commenters reported nearly identical associations, for example, Chapter 2, example (9) reprinted below as (15).

(15) Depends... Idk why but in my head, passed away = natural causes/illness/age

Died = anything else/accident/unexpected

Why this is the case may be closely linked to implications of the metaphoric structure of *PASS*-type variants.

First, *DIE* and *PASS*-type variants may not share identical lexical aspect. Because *DIE* describes the moment of transition from the state of being alive to being dead, it is often given as a prototypical example of a punctual telic verb (Dowty, 1991, p. 68; Rothstein, 2004, pp. 22–23; Vendler, 1957). The inherent endpoint of such verbs is nearly instantaneous and so they are said to have no internal structure (Rothstein, 2004, p. 22). Though *PASS*-type variants can describe the same punctual event, unlike *DIE*, these variants *do* have internal structure. As DEATH IS DEPARTURE metaphors, *PASS*-type variants are connected to path motion, the metaphor’s source domain (Sullivan & Wachowski, 2020). As a result, there may be an aspectual mismatch between the abrupt nature of violent/sudden death and the durative nature of the metaphorical motion underlying *PASS*-type variants.

Second, *DIE* takes an experiencer subject and is non-volitional without additional context (e.g. *she chose to die on her own terms*). In contrast, animate subjects of journey verbs¹¹ are typically agents (e.g. *she crossed the dance floor; they left, we went home*) and such movements appear volitional without additional context (e.g. *they made her cross the dance floor*). While *PASS*-type variants are not truly agentive, evoking a journey attributes more agency to the subject than *DIE* allows. For this reason, the violation of agency in wrongful death is particularly at odds with *pass (away/on)*, and peaceful deaths, where the deceased are “ready” to die for example, more acceptable.

¹¹ This is certainly not true of *all* verbs of path motion. E.g. *she tumbled down the stairs; they fell off the dock*.

Third, when a person is said to *PASS (away/on)*, what is conceptualized as moving from SOURCE to GOAL? Comments analyzed in Chapter 2, section 2.4 suggest that it is not (at least principally) the deceased’s physical body, but rather who they are as a person, whether or not this metaphysical selfhood is defined as an immortal soul or spirit. Perhaps for many, *PASS (away/on)* is a spiritual or metaphysical event that strongly implies bodily death, while *DIE* explicitly describes the death of the body and implies the departure of the self. Acknowledging a cause of death means discussing its physiological details, and perhaps using the variant *DIE* is consonant with this bodily focus. That tokens of unknown DEATH TYPE use the highest proportion of *PASS*-type variants is consistent with this hypothesis. An alternative (or additional) explanation is that discussing cause of death, particularly when violent, is likely about conveying information rather than providing support and, as shown elsewhere, informative communication favours direct language while supportive communication favours euphemism.

PASS-type variants also become more likely when the speaker has a PERSONAL RELATIONSHIP with the deceased, for at least three possible reasons. First, indirect language can serve a distancing function from the emotional discomfort or grief tied to the deceased’s absence (as described in Chapter 2, section 2.5). Second, speakers may more readily conceptualize the continued existence of the dead they have known and evoke this possibility with DEATH IS DEPARTURE metaphors. Finally, qualitative findings (Chapter 2, section 2.6) suggest that North Americans find *PASS*-type variants more polite than direct language, which may motivate their interpersonal use. These euphemistic variants treat a taboo topic indirectly, which to Brown and Levinson (1999, p. 314) is a form of positive politeness, since it shows that the speaker values the addressee’s perspective and emotions. Use of *PASS*-type variants could also be framed as negative politeness, mitigating the face threatening act of emotional imposition on the addressee

(Leech, 2014, pp. 25–26). Both framings, however, concern impacts on an addressee, yet it is the relationship between the speaker and the deceased that conditions variant choice in the dataset.

Two additional considerations allow us to see the connection. 1) Drawing on Bell's (1984) Audience Design framework, the deceased can be thought of as a referee when death is reported, an interactional participant who, though not present, is as salient to style as the addressee (Ford & Loureiro-Rodríguez, forthcoming). 2) Though the politeness function of *PASS*-type variants likely has its root in active face-threat mitigation as described above, use of these variants has likely conventionalized, what Leech (2014, p. 14) calls *pragmaticalization*, rendering them conventionally (i.e. pragmalinguistically; see Figure 3) polite. Thus, with politeness in mind, speakers may choose these variants to show respect for people they have known.

Though both of the highly correlated potential predictors *EVENT TIME* and *DEATH HAS OCCURRED* had positive variable importance, logistic regression establishes that it is the latter that truly matters. Deaths that have occurred required more softening with euphemistic *PASS*-type variants than those that haven't, likely because of their potential emotional impact on both the speaker and addressee (Chapter 2, sections 2.5 and 2.6; Brown & Levinson, 1999, p. 314). “Indeterminate” tokens, in contrast, favoured *DIE*. In addition to being difficult to classify regarding whether death has occurred, what unites many of these tokens is a subject with non-specific reference, as illustrated by examples (16-18). Even if, in a sense, these examples discuss real death, lack of specificity makes them highly abstract and impersonal. Without even the possibility of a *PERSONAL RELATIONSHIP* with the deceased, the interpersonal motivations to use *PASS*-type variants discussed above cannot apply.

(16) [...] of course, people die while they're waiting. (CBCA 2019)

- (17) This may be the only place where this name appears because when a trans person dies, the body reclaimed by the family of origin and their name then evaporates, so they die twice. (CBCA 2019)
- (18) You can still present all the evidence that the person has died, but that often becomes the defense is you can't even prove that the person is dead and that becomes something that the prosecution has to overcome. (COCA:SPOKEN 2019)

Other social factors investigated were not significant predictors of variant choice in logistic regression models. First, it is possible that the focus on conveying information and frequent discussion of sudden/violent death on broadcast media so strongly influence variant choice toward *DIE*, that weaker effects of factors hypothesized to favour *PASS*-type variants are drowned out and can't be examined in this context. For example, based on chapter 2 sections 2.5 and 2.6 I would have expected speakers to use more *PASS*-type variants when supporting recently bereaved interlocutors. In the dataset, however, this situation rarely occurred (n=29/1000) and when it did, the deaths involved were usually “remarkable” (72%). Lack of effect of *SPEAKER GENDER* is not unexpected, since *DIE* and *PASS away* (which accounts for the majority of *PASS*-type variants) are longstanding variants that appear to be in stable variation in recent decades (Ford & Loureiro-Rodríguez, forthcoming). Had it been possible to examine factors conditioning the use of *PASS*, believed to be an incoming variant, we may have seen a gender effect. Lack of effect of *DECEASED GENDER* is consistent with Ford and Loureiro-Rodríguez (forthcoming) but contrary to Ferraro (2019), adding a data point to this open question.

Results for *STATUS AS JOURNALIST* were inconsistent between analyses: it had high variable importance in random forest, but was non-significant in logistic regression. This may be explained by covariance between *STATUS AS JOURNALIST* and multiple significant predictors.

Within the dataset, journalists report the sudden or violent death of people they don’t know personally more often than non-journalists (54% versus 31%). This makes JOURNALIST a partial proxy for the significant predictors DEATH TYPE and RELATION TO SPEAKER. Since each tree in the random forest is built using five randomly selected potential predictors, when JOURNALIST is selected, but DEATH TYPE and RELATION TO SPEAKER are not, JOURNALIST stands in for them (e.g. Figure 8b). In addition, the influence of atypical speakers who contributed a large number of tokens may have had an outsized impact on random forest that was neutralized in logistic regression by limiting data for analysis to one-observation-per-speaker.

Ethnicity, a social factor that was not investigated here may also play a role in variant choice. In the Canadian data, four out of five tokens of *PASS on* collected were contributed by Indigenous speakers. Nothing can be concluded from this due to the very small numbers involved, but it suggests that the role of ethnicity or related communities of practice in variant choice should be investigated in the future.

3.7.3 Does use of variants in this context differ according to linguistic factors?

Among potential linguistic predictor variables, use of the PROGRESSIVE shows the greatest evidence of influence on variant choice. It had high variable importance in random forest analysis, improved logistic regression models and showed a trend toward significance. If this trend is taken at face value, why might *DIE* be more acceptable in the progressive than *PASS*-type variants?

First, whether the progressive¹² is available and what it means differs between verbs of different aspectual classes. As discussed in section 3.7.2, *DIE* is considered a prototypical

¹² When referring to the predictor variable, PROGRESSIVE appears in small capitals. When referring to progressive aspect itself/its morphosyntactic marking, I’ve used regular typeface.

punctual telic verb with no internal structure (Dowty, 1991, p. 68; Rothstein, 2004, pp. 22–23; Vendler, 1957). The progressive, however, can introduce duration before the telic point when a punctual telic verb has associated preparatory events that lead to its culmination (Rothstein, 2004, p. 52). For example, though no death has yet occurred in (19), the progressive indicates that associated physiological or mental changes—i.e. bodily processes—that ultimately lead to death are underway. This contrasts with durative telic events like (20), where the progressive implies that the event itself is underway and some increment of it accomplished.

(19) Her grandfather is dying. (If true, has her grandfather died? No)

(20) Bayan is walking to the store. (If true, has Bayan walked? Yes)

Unlike *DIE*, as DEATH IS DEPARTURE metaphors, *PASS*-type variants are structured on the SOURCE-PATH-GOAL image schema that defines path motion (Sullivan & Wachowski, 2020). Just like *Bayan* in example (20), the deceased travels from an implied SOURCE to some other location, albeit metaphorically, giving the event a metaphorical, if not literal, duration. I would suggest that, as a result, use of the progressive in (21) implies that the journey is underway, just like in (20).

(21) ?#Her grandfather is passing away.

For this reason, compared to *DIE*, the use of *PASS*-type variants in the progressive may be limited less by grammatical acceptability, and more by a narrower context of use. While (19) is appropriate even if death is not expected for weeks or months, (21) is most felicitous if *her grandfather* is in the moment of death.

Second, even if *PASS-type variants* behave as punctual telic verbs in most respects, their metaphorical structure may limit the application of the progressive by introducing a focus on the

departure of the self rather than physical death. We saw earlier that the progressive can introduce duration before the telic point when a punctual verb has associated preparatory events that lead to its culmination (Rothstein, 2004, p. 52). For *DIE* these are the bodily changes that will lead to death. Perhaps *PASS*-type variants lack metaphysical analogues to the physical changes that occur before the death of the body, making use of the progressive in any sense but “currently in progress” impossible. If true, this would provide an additional explanation for why use of progressive *PASS*-type variants appears limited to the moment of death and therefore less common than *DIE* in the progressive. In sum, the structure of *PASS*-type variants as DEATH IS DEPARTURE metaphors may limit their use in the progressive, both because of the aspectual properties implied by PATH-motion and the focus on metaphysical, rather than bodily, death.

Other linguistic factors investigated (main or embedded CLAUSE, modals by various measures, PRESENT and PAST PERFECT) were not significant predictors of variant choice in logistic regression models. PRESENT PERFECT and CLAUSE did have positive conditional variable importance in random forest analysis, but there are several possible factors that could account for this. First, the magnitude of any overall effect of these predictors may not be large enough to detect in a regression model. Second, effects may be due to atypical speakers as discussed in section 3.7.2. Finally, these predictors may only affect variant choice in very specific situations. For example, the conditional inference tree in Figure 9 indicates that, in this dataset, when talking about the sudden or violent death of a friend, stranger or oneself, use of the present perfect favours more frequent use of *PASS*-type variants. Interactions of this complexity cannot be modeled by logistic regression on a dataset of this size (they would not converge). It is possible that, in this situation, highlighting a violent or sudden death (that would generally favour use of *DIE*) with the present perfect triggers compensatory “softening” with euphemistic variants. For

example, in (22) *the ones who have already passed on* are people who died by suicide and softening a highlighted taboo topic in this way avoids threatening the addressee’s positive face by showing care for their emotions (Brown & Levinson, 1999, p. 314).

(22) We can't help the ones that have already passed on, but help these ones that are seeking help.

Nevertheless, the specificity of the lower splits of CIT analysis may reflect overfitting to this particular dataset, and not be generalizable beyond it.

3.7.4 Does use of variants differ between Canada and the United States?

In the standard, national varieties represented in broadcast media, Canadians and Americans choose between *DIE* and *PASS*-type variants at similar rates. Much like STATUS AS JOURNALIST, COUNTRY showed high variable importance in random forest analysis, but was not a significant predictor of the choice between *PASS*-type variants in logistic regression. Unlike JOURNALIST, it is not an obvious proxy for a significant predictor. It seems likely that its high variable importance is related to complex interactions in the random forest—low-level splits across many trees reflecting differences in use in very specific situations. Figure 8a illustrates this. In it, Canadian speakers use *DIE* more often than Americans in one situation, but less in another. Figure 8a is only one of 1000 trees in the random forest, however, and overall CIT analysis agrees with logistic regression, showing no effect of COUNTRY, interaction or otherwise.

3.8 Conclusion

Use of variants of ‘DIE’ in spontaneous Canadian and American broadcast speech is strongly influenced by the informative focus of the medium toward the direct language variant *DIE*. Nevertheless, variation is present within this context and conditioned by social factors (DEATH TYPE, PERSONAL RELATIONSHIP and whether DEATH HAS OCCURRED) and possibly

linguistic factors as well (use of the PROGRESSIVE). Deaths that were sudden or violent are most often reported with the punctual, bodily-focused *DIE*, while “unremarkable” deaths due to chronic illness or unknown causes use *PASS*-type variants more often. *PASS*-type variants are most common, however, when interpersonal factors come into play. When the death of someone personally known to the speaker has occurred, speakers choose *PASS*-type variants more often, perhaps to gentle grief or discomfort, to evoke the possibility that those they have known continue to exist, or to show the deceased respect. Use of the PROGRESSIVE may predict *DIE*, since *DIE* occurs freely in the progressive with a range of meanings, whereas *PASS*-type variants appear limited to ‘moment of death’ interpretations because of the implications of their metaphoric structure. COUNTRY was not a significant predictor of the choice between *DIE* and *PASS*-type variants, suggesting that if there are differences in how Canadians and Americans use variants of ‘DIE,’ they are not represented in the standard national varieties favoured by broadcast media.

Using multiple statistical methods (random forest, CIT analysis, logistic regression) presented both challenges and benefits. The principal challenge was interpretation of conflicting results, for example when random forest identified a variable as highly important, but it was not a significant predictor in logistic regression. At the same time, what such conflicts offer is a more nuanced view of the data. For example, the discrepancy between the performance of CLAUSE and PRESENT PERFECT in random forest and logistic regression with one observation per speaker revealed the potential influence of atypical individual speakers on random forest. Both tree methods also helped guide regression analysis in different ways: variable importance with random forest helped prioritize predictors for model building, while CIT analysis guided grouping of multi-leveled predictors.

This study provides a window into use of variants of ‘DIE’ in a particular communicative context. The nature of that context and its investigation through general and “non-conventional” corpora, bring certain limitations. The overwhelming preference for direct language variants in broadcast media speech precluded examination of individual *PASS*-type variants and the factors that might condition their use. As spoken data, broadcast speech is also unusually skewed toward formal standard language, which may have limited my ability to learn more about *PASS*, a variant that is potentially less formal in Canada (Ford & Loureiro-Rodríguez, forthcoming) and regional in the US (Cassidy & Hall, 1985-). The context is also relationally removed from everyday life: interlocutors did not typically have an ongoing social role in each other’s lives, nor a shared connection to the deceased, and discussed the “unremarkable” sort of deaths that touch the lives of most people less often than those whose atypicality made them newsworthy. Drawing data from corpora that are not annotated for social factors like social class, ethnicity, and age, also means that any role these factors play in variation could not be taken into account, nor could I look at change in apparent time.

Future research might examine variation in contexts that skew less toward informative communication and standard language, like sociolinguistic interviews or casual conversation between participants who know one other. Such work might explicitly test the role of region, ethnicity and register on variation, and answer questions that could not be resolved in this inquiry. For example, does earlier attestation in the US (1950s, interview) than Canada (2000s, obituaries) mean that *PASS* originates in the US, or has parallel Canadian use simply not been noted in the literature? Looking at use of variants over real or apparent time in a common context could shed light on this. DARE (1985-) asserts that *PASS* is a regional variant used in the American South, and particularly by Black speakers, but we don’t know whether it is enregistered as such

in either the US or Canada. If an association with Southern American English and African American Language—both non-standard, stigmatized varieties (Lippi-Green, 2012)—exists, could this affect use in relatively formal registers (like broadcast media) in either country? Similar questions about the role of *PASS on* and whether it is truly used more frequently by Indigenous speakers in Canada could also be investigated in future work.

Chapter 4: Conclusion

This thesis examined variation in ‘DIE’ to understand how lexical choices reflect societal perceptions of a universal human experience. Using a mixed-methods approach made it possible not only to identify the factors shaping this variation but also to understand the meanings attached to it.

Qualitative investigation guided the design of the quantitative study in several ways. First, Reddit commenters highlighted the shared denotative meaning of variants, supporting treatment of ‘DIE’ as a lexical variable. They also confirmed the relevance of *PASS* (*away/on*) beyond medical and obituary contexts and identified the divide between direct language and euphemism as the most salient distinction between variants, which supported treatment of *PASS*-type variants as a group. In addition, qualitative themes such as manner of death (DEATH TYPE), experience of grief (DEATH RECENCY), and regional variation (COUNTRY) informed coding of the quantitative analysis.

The clearest point of convergence between Chapters 2 and 3 was the relationship between manner of death and variant choice. Commenters identified this as a hunch, or mental association between *DIE* and sudden or violent deaths on one hand, and *PASS*-variants and gradual deaths on the other. Quantitative analysis saw this borne out in usage. Commenters also claimed that they used *DIE* when the deceased was not known to either interlocutor, and quantitative analysis validated its inverse, that *PASS*-type variants are more likely when even the speaker alone has a personal relationship with the deceased.

In addition to mutually reinforcing each other’s findings, the two studies explained and elaborate each other’s results. For example, the qualitative finding that a focus on communicating information favours direct language helped explain the high overall frequency of

DIE relative to *PASS*-type variants in the broadcast media dataset. At the same time, quantitative results showed that the influence of a communicative context focused on information is not limited to reporting death in professional healthcare and journalism roles (the focus of the qualitative discussion), since both journalists and non-journalists used high rates of *DIE*.

Perhaps one of the most interesting threads woven through both studies is the influence of the metaphorical nature of *PASS*-type variants on their use relative to *DIE*. Their metaphysical, rather than bodily focus as DEATH IS DEPARTURE metaphors and euphemistic indirectness motivate interpersonal functions related to emotional discomfort, grief and politeness. The metaphoric structure of *pass*-type variants is no less important, creating subtle thematic-role and aspectual contrasts with *DIE* that connect to social predictors of variation like manner of death, and possibly even linguistic factors like the PROGRESSIVE, shedding light on a previously unexamined facet of lexical variation.

Taken together these two studies characterize *DIE* as informative, bodily focused, and impersonal and *PASS away*, *PASS on* and *PASS* as specialized variants recruited for relationship-focused communication. Future work may look at how these domain-specific roles developed over time, and investigate the social meaning attached to individual *PASS*-type variants.

References

- Adams, M. (2014). Vocabulary analysis in sociolinguistic research. In J. Holmes & K. Hazen (Eds.), *Research methods in sociolinguistics: A practical guide* (pp. 163–193). Wiley Blackwell.
- Afful, J. B. (2012). A genre analysis of death announcements in Ghanaian newspapers. *Language Discourse & Society*, 1(2), 118–137.
- Aijmer, K. (2018). 'That's well bad: Some new intensifiers in spoken British English. In V. Brezina, R. Love, & K. Aijmer (Eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the spoken BNC2014* (1st ed., pp. 60–95). Routledge.
<https://doi.org/10.4324/9781315268323-6>
- Allan, K., & Burridge, K. (1991). *Euphemism & dysphemism: Language used as shield and weapon*. Oxford University Press.
- Aries, P. (1974). The Reversal of death: Changes in attitudes toward death in Western societies. *American Quarterly*, 26(5), 536–560. <https://doi.org/10.2307/2711889>
- Arnup, K. (2018). *Family perspectives: Death and dying in Canada*. The Vanier Institute of the Family.
- Baker, P. (2010a). *Sociolinguistics and corpus linguistics*. Edinburgh University Press.
- Baker, P. (2010b). Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language*, 4(1), 125–149.
- Barber, K. (2004a). euphemism. In *The Canadian Oxford dictionary* (2nd ed.). Oxford University Press.
- Barber, K. (2004b). pass. In *The Canadian Oxford dictionary* (2nd ed.). Oxford University Press.

- Barlet, M. H., Barks, M. C., Ubel, P. A., Davis, J. K., Pollak, K. I., Kaye, E. C., Weinfurt, K. P., & Lemmon, M. E. (2022). Characterizing the language used to discuss death in family meetings for critically ill infants. *JAMA Network Open*, 5(10), E2233722.
<https://doi.org/10.1001/jamanetworkopen.2022.33722>
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204.
<https://doi.org/10.1017/S004740450001037X>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brezina, V., & Meyerhoff, M. (2014). Significant or random?: A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics*, 19(1), 1–28. <https://doi.org/10.1075/ijcl.19.1.01bre>
- Brook, M., Blamire, E., & Tagliamonte, S. A. (2025). A linguistic phoenix: The recycling of very in Ontario English. *Language Variation and Change*, 37(2), 191–211.
<https://doi.org/10.1017/S0954394525100525>
- Brown, P., & Levinson, S. C. (1999). Politeness: Some universals in language useage. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (2nd ed., pp. 311–323). Routledge.
- Buckle, R. (1980). *U and non-u revisited* (1st ed.). Debrett's Peerage Ltd. in association with Futura Publications.
- Campbell-Kibler, K. (2010). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change*, 22(3), 423–441. <https://doi.org/10.1017/S0954394510000177>
- Cassidy, F. G., & Hall, J. H. (1985). pass v. In *Dictionary of American regional English, Volume III*. Belknap Press of Harvard University Press.
- Cebrat, G. (2016). *Death notice as a genre : An analysis of the New York Times online edition*.

Doctoral dissertation, Uniwersytet Śląski.

Cheshire, J. (2007). Discourse variation, grammaticalisation and stuff like that. *Journal of*

Sociolinguistics, 11(2), 155–193. <https://doi.org/10.1111/j.1467-9841.2007.00317.x>

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, 66(4), 764–805.

<https://doi.org/10.2307/414729>

Crespo Fernández, E. (2006). The language of death: Euphemism and conceptual

metaphorization in Victorian obituaries. *SKY Journal of Linguistics*, 19, 101–130.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*.

<https://www.english-corpora.org/coca/>

Denis, D. (2021). Raptors vs. Bucktees: The Somali influence on Toronto slang. *Journal of*

Multilingual and Multicultural Development, 42(6), 565–578.

<https://doi.org/10.1080/01434632.2021.1895181>

Dines, E. R. (1980). Variation in discourse—“and stuff like that.” *Language in Society*, 9(1), 13–

31. <https://doi.org/10.1017/S0047404500007764>

Dowty, D. R. (1991). *Word meaning and Montague grammar : The semantics of verbs and times*

in generative semantics and in Montague's PTQ (Reprinted). Kluwer Academic.

Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics*,

28(2), 105–144. <https://doi.org/10.1162/089120102760173625>

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora : A*

practical framework for corpus representativeness (1st ed.). Cambridge University Press.

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in

social media. *PloS One*, 9(11), e113114. <https://doi.org/10.1371/journal.pone.0113114>

Ferraro, F. R. (2019). Males tend to die, females tend to pass away. *Death Studies*, 43(10), 665–

667. <https://doi.org/10.1080/07481187.2018.1515127>

Ford, L. (2026). *Lives, time and emotion in motion: A diachronic corpus analysis of passed away*. North American Cognitive Linguistics Association Conference.

Ford, L., & Loureiro-Rodríguez, V. (forthcoming). From *died* to *passed*: Lexical variation and change in Canadian newspaper obituaries. *Canadian Journal of Linguistics*.

Franco, K., & Tagliamonte, S. A. (2021). Interesting fellow or tough old bird? Third-person male referents in Ontario. *American Speech*, 96(2), 192–216. <https://doi.org/10.1215/00031283-8661833>

Froke, P., Bratton, A. J., Gross, A., McMillan, J., Sarkar, P., & Schwartz, J. (2024). *The Associated Press Stylebook* (57th ed.). The Associated Press.

Gardner, M. H., Denis, D., Brook, M., & Tagliamonte, S. A. (2021). Be like and the Constant Rate Effect: from the bottom to the top of the S-curve. *English Language and Linguistics*, 25(2), 281–324. <https://doi.org/10.1017/S1360674320000076>

Goodman, N. (1949). On likeness of meaning. *Analysis (Oxford)*, 10(1), 1–7. <https://doi.org/10.1093/analys/10.1.1>

Gorer, G. (1955). The pornography of death. *Encounter, October*, 49–52.

Green, L. (2000). Aspectual be—type constructions and coercion in African American English. *Natural Language Semantics*, 8(1), 1–25. <https://doi.org/10.1023/A:1008392112529>

Greenwell, B. M. (2022). *Tree-based methods : A practical introduction with applications in R*. CRC Press.

Grieve, J., Nini, A., & Guo, D. (2018). Mapping lexical innovation on American social media. *Journal of English Linguistics*, 46(4), 293–319. <https://doi.org/10.1177/0075424218793191>

Habeck-Fardy, A. (2021). Passing on death: An audit of the terminology utilized in discharge

summaries for deceased patients. *Death Studies*, 45(6), 413–419.

<https://doi.org/10.1080/07481187.2019.1648335>

Hansen, B. (2018). *Corpus linguistics and sociolinguistics: A study of variation and change in the modal systems of world Englishes* (1st ed.). Brill.

Harper, D. (n.d.). *pass out* (v.). Online Etymology Dictionary.

[https://www.etymonline.com/word/pass out](https://www.etymonline.com/word/pass%20out)

Henderson Taylor, S. (1974). Terms for low intelligence. *American Speech*, 49(3/4), 197–207.

<https://doi.org/10.2307/3087798>

Holt, E. (1996). Reporting on talk: The use of direct reported speech in conversation. *Research on Language and Social Interaction*, 29(3), 219–245.

https://doi.org/10.1207/s15327973rlsi2903_2

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674.

<https://doi.org/10.1198/106186006X133933>

Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, 16, 3905–3909.

Ito, R., & Tagliamonte, S. A. (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers. *Language in Society*, 32(2), 257–279.

<https://doi.org/10.1017/S0047404503322055>

Jankowski, B. L., & Tagliamonte, S. A. (2019). Supper or dinner? Sociolinguistic variation in the meals of the day. *English World-Wide*, 40(2), 169–200.

<https://doi.org/10.1075/eww.00027.jan>

Jankowski, B. L., & Tagliamonte, S. A. (2021). Urban-rural dimensions to variable -body/-one.

- In G. Oberdorfer, S. Edler, & A. Ziegler (Eds.), *Urban Matters: Current approaches in variationist sociolinguistics* (pp. 141–158). John Benjamins Publishing Company.
<https://doi.org/10.1075/silv.27>
- Johnson, E. (1996). *Lexical change and variation in the Southeastern United States, 1930-1990* (1st ed.). University of Alabama Press.
- Johnston, T., Cresdee, D., Schembri, A., & Woll, B. (2015). FINISH variation and grammaticalization in a signed language: How far down this well-trodden pathway is Auslan (Australian Sign Language)? *Language Variation and Change*, 27(1), 117–155.
<https://doi.org/10.1017/S0954394514000209>
- Kany, C. E. (1960). *American-Spanish euphemisms*. University of California Press.
- Kendall, T. (2011). Corpora from a sociolinguistic perspective. *Revista Brasileira de Linguística Aplicada*, 11(2), 361–389. <https://doi.org/10.1590/s1984-63982011000200005>
- Kiesling, S. F. (2004). Dude. *American Speech*, 79(3), 281–305.
<https://doi.org/10.1215/00031283-79-3-281>
- Kircher, R. (2022). Analysis of the societal treatment of language. In *Research methods in language attitudes* (pp. 17–96). Cambridge University Press.
<https://doi.org/10.1017/9781108867788.003>
- Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language (Baltimore)*, 45(4), 715–762. <https://doi.org/10.2307/412333>
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Labov, W. (1982). Building on empirical foundations. In W. P. Lehmann & Y. Malkiel (Eds.), *Perspectives on historical linguistics* (1st ed., pp. 17–92). J. Benjamins.
<https://doi.org/10.1075/cilt.24>

- Labov, W. (2006). *The social stratification of English in New York City* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511618208>
- Labov, W. (2012). *Dialect diversity in America: The politics of language change* (1st ed.). University of Virginia Press.
- Leech, G. (2014). *The pragmatics of politeness* (1st ed.). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195341386.001.0001>
- Lehmann, H. M. (2002). Zero subject relative constructions in American and British English. In P. Peters (Ed.), *New Frontiers of Corpus Research* (Issue 36, pp. 163–177). Editions Rodopi.
- Lindsey, G. (2019). What was RP? In G. Lindsey (Ed.), *English after RP: Standard British pronunciation today* (pp. 1–5). Springer International Publishing. https://doi.org/10.1007/978-3-030-04357-5_1
- Lippi-Green, R. (2012). *English with an accent : Language, ideology, and discrimination in the United States*. Routledge.
- Matiki, A. (2001). The pragmatics of obituary notices in Malawian newspapers. *Language Matters (Pretoria, South Africa)*, 32(1), 29–52. <https://doi.org/10.1080/10228190108566171>
- McEnery, T., & Brezina, V. (2022). *The fundamental principles of corpus linguistics* (1st ed.). Cambridge University Press.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics : Method, theory and practice*. Cambridge University Press.
- Mitford, N., Ross, A. S. C., Waugh, E., Sykes, C., & Betjeman, J. (1956). *Noblesse oblige: An enquiry into the identifiable characteristics of the English aristocracy* (N. Mitford (ed.)). H.

Hamilton.

Moran, C. (2003). Beyond content: Does humor help coping? *Disability Studies Quarterly*, 23(3/4). <https://doi.org/10.18061/dsq.v23i3/4.434>

Morin, C., & Grieve, J. (2024). The semantics, sociolinguistics, and origins of double modals in American English: New insights from social media. *PloS One*, 19(1), e0295799. <https://doi.org/10.1371/journal.pone.0295799>

Needle, J. M., & Tagliamonte, S. A. (2025). Buddies, dudes, and bros of Ontario: Trends and patterns of vocative change. *Journal of English Linguistics*, 53(2), 107–136. <https://doi.org/10.1177/00754242251341318>

O’Sullivan, J., Amador-Moreno, C. P., & Barron, A. (2025). *Corpus linguistics for sociolinguistics: A guide for research* (1st ed.). Taylor & Francis Group. <https://doi.org/10.4324/9780429300325>

Ondimu, J. (2014). A socio-cultural understanding of death: A genre analysis of obituaries in a Kenyan newspaper. *Language Matters (Pretoria, South Africa)*, 45(1), 3–22. <https://doi.org/10.1080/10228195.2013.805799>

Oushiro, L. (2023). Computational resources for handling sociolinguistic corpora. In M. Díaz-Campos & S. Balasch (Eds.), *The handbook of usage-based linguistics variation, change, and the usage-based approach* (pp. 417–434). <https://doi.org/https://doi-org.uml.idm.oclc.org/10.1002/9781119839859.ch23>

Oxford University Press. (n.d.-a). die, v.¹, sense I.1.a. In *Oxford English dictionary*. Oxford University Press. <https://doi.org/10.1093/OED/1583811001>

Oxford University Press. (n.d.-b). pass, v. In *Oxford English dictionary*. Oxford University Press. <https://doi.org/https://doi.org/10.1093/OED/1206119682>

- Phillips, J. B. (2007). The changing presentation of death in the obituary, 1899–1999. *Omega: Journal of Death and Dying*, 55(4), 325–346. <https://doi.org/10.2190/OM.55.4.g>
- ProQuest. (n.d.). *About Canadian Business & Current Affairs database*. Retrieved April 7, 2026, from <https://www.proquest.com/cbcacomplete/productfulldescdetail>
- R Core Team (2023). *R: A Language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
- Rawlings, D., Tieman, J. J., Sanderson, C., Parker, D., & Miller-Lewis, L. (2017). Never say die: Death euphemisms, misunderstandings and their implications for practice. *International Journal of Palliative Nursing*, 23(7), 324–330. <https://doi.org/10.12968/ijpn.2017.23.7.324>
- Rayson, P., Leech, G. N., & Hodges, M. (1997). Social differentiation in the use of English vocabulary: Some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics*, 2(1), 133–152. <https://doi.org/10.1075/ijcl.2.1.07ray>
- Ross, A. S. C. (1954). Linguistic class-indicators in present-day English. *Neuphilologische Mitteilungen*, 108(1a), 113–149.
- Rothstein, S. D. (2004). *Structuring events: A study in the semantics of lexical aspect*. Blackwell.
- Säily, T. (2011). Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory*, 7(1), 119–141. <https://doi.org/10.1515/cllt.2011.006>
- Sallnow, L., Smith, R., Ahmedzai, S. H., Bhadelia, A., Chamberlain, C., Cong, Y., Doble, B., Dullie, L., Durie, R., Finkelstein, E. A., Guglani, S., Hodson, M., Husebø, B. S., Kellehear, A., Kitzinger, C., Knaul, F. M., Murray, S. A., Neuberger, J., O'Mahony, S., ... Wyatt, K. (2022). Report of the Lancet Commission on the value of death: Bringing death back into

- life. *The Lancet*, 399(10327), 837–884. [https://doi.org/https://doi.org/10.1016/S0140-6736\(21\)02314-X](https://doi.org/https://doi.org/10.1016/S0140-6736(21)02314-X)
- Sadow, R. (2024). The role of local identity in the usage and recognition of Anglo-Cornish dialect lexis. *Journal of English Linguistics*, 52(1), 38–64. <https://doi.org/10.1177/00754242231220343>
- Sadow, R., Grieve, J., Stamp, R., & Schembri, A. (2026). Lexical variation and change: Integrating lexis into variationist sociolinguistics. *Language Variation and Change*, 37(3), 293–318. <https://doi.org/10.1017/S0954394525100628>
- Sankoff, D., & Thibault, P. (1981). Weak complementarity: Tense and aspect in Montreal French. In B. B. Johns & D. R. Strong (Eds.), *Syntactic change*. University of Michigan Press.
- Schmid, H.-J. (2003). Do men and women really live in different cultures? Evidence from the BNC. In A. Wilson, R. Rayson, & T. McEnerey (Eds.), *Corpus Linguistics by the Lune* (pp. 185–221). Peter Lang.
- Sigley, R., & Holmes, J. (2002). Looking at girls in corpora of English. *Journal of English Linguistics*, 30(2), 138–157. <https://doi.org/10.1177/007242030002004>
- Smithson, A.-M. (2019). *Discourses in stone: Dialogues with the dissenting dead 1830-1919*. Doctoral dissertation, University of Chester.
- Solheim, M. A. D. (2014). *Death by metaphor: A study of metaphors and conceptualisations of death in British and American obituaries*. Doctoral dissertation, University of Oslo.
- Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., & Cormier, K. (2014). Lexical variation and change in British Sign Language. *PloS One*, 9(4), e94053. <https://doi.org/10.1371/journal.pone.0094053>

- Stratton, J. M. (2022). Tapping into German adjective variation: A variationist sociolinguistic approach. *Journal of Germanic Linguistics*, 34(1), 63–102.
<https://doi.org/10.1017/S1470542721000088>
- Stratton, J. M. (2023). Where did *wer* go? Lexical variation and change in third-person male adult noun referents in Old and Middle English. *Language Variation and Change*, 35(2), 199–221. <https://doi.org/10.1017/S0954394523000108>
- Stratton, J. M., & Sundquist, J. D. (2022). A variationist sociolinguistic analysis of intensifiers in Oslo Norwegian. *Journal of Germanic Linguistics*, 34(4), 385–419.
<https://doi.org/10.1017/S1470542722000022>
- Strick, M., Holland, R. W., van Baaren, R. B., & van Knippenberg, A. (2009). Finding comfort in a joke: Consolatory effects of humor through cognitive distraction. *Emotion (Washington, D.C.)*, 9(4), 574–578. <https://doi.org/10.1037/a0015951>
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307.
<https://doi.org/10.1186/1471-2105-9-307>
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25.
<https://doi.org/10.1186/1471-2105-8-25>
- Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on!—A new, conditional variable importance measure for random forests available in party. *The R Journal*, 1(2), 14.
<https://doi.org/10.32614/RJ-2009-013>
- Sullivan, K., & Wachowski, W. (2020). Everyone “leaves” the world eventually: Culture-based homogeneity and variation in death is departure. *Review of Cognitive Linguistics*, 18(1), 75–

93. <https://doi.org/10.1075/rcl.00051.sul>

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English:

Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–178. <https://doi.org/10.1017/S0954394512000129>

Tagliamonte, S. A., & Brooke, J. (2014). A weird (language) tale: Variation and change in the adjectives of strangeness. *American Speech*, 89(1), 4–41. <https://doi.org/10.1215/00031283-2726386>

Tagliamonte, S. A., & D’Arcy, A. (2004). He’s like, she’s like: The quotative system in Canadian youth. *Journal of Sociolinguistics*, 8(4), 493–514. <https://doi.org/10.1111/j.1467-9841.2004.00271.x>

Tagliamonte, S. A., & Jankowski, B. L. (2019). Golly, gosh, and oh my God! What North American dialects can tell us about swear words. *American Speech*, 94(2), 195–222. <https://doi.org/10.1215/00031283-7251241>

Tagliamonte, S. A., & Jankowski, B. L. (2025). Lexical variation of woods and bush in Ontario English. *World Englishes*, 44(1–2), 285–299. <https://doi.org/10.1111/weng.12704>

Tagliamonte, S. A., & Pabst, K. (2020). A cool comparison: Adjectives of positive evaluation in Toronto, Canada and York, England. *Journal of English Linguistics*, 48(1), 3–30. <https://doi.org/10.1177/0075424219881487>

Tagliamonte, S. A., & Roberts, C. (2005). So weird; so cool; so innovative: The use of intensifiers in the television series Friends. *American Speech*, 80(3), 280–300. <https://doi.org/10.1215/00031283-80-3-280>

Tagliamonte, S. A., & Smith, J. (2021). Obviously undergoing change: Adverbs of evidentiality across time and space. *Language Variation and Change*, 33(1), 81–105.

<https://doi.org/10.1017/S0954394520000216>

Trudgill, P. (1986). *Dialects in Contact*. Blackwell.

Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, 66(2), 143–160.

<https://doi.org/10.2307/2182371>

Walter, T. (1994). Traditional, modern and neo modern death. In *The revival of death* (pp. 194–209). Routledge. <https://doi.org/10.4324/9780203220306-18>

Williams, A., Kerswill, P., Foulkes, P., & Docherty, G. J. (1999). Dialect levelling: Change and continuity in Milton Keynes, Reading and Hull. In *Urban voices* (1st ed., pp. 141–162). Routledge. <https://doi.org/10.4324/9781315832470-8>

Wilson, E., Caswell, G., Turner, N., Pollock, K., Dunn, A. M., Hofmann, O. S., Waters, B., & Witchel, E. (2024). Talking about death and dying: Findings from deliberative discussion groups with members of the public. *Mortality (Abingdon, England)*, 29(1), 176–192. <https://doi.org/10.1080/13576275.2022.2136515>

Wolfram, W. (1974). The relationship of white Southern speech to Vernacular Black English. *Language (Baltimore)*, 50(3), 498–527. <https://doi.org/10.2307/412221>

Wolfram, W. (1993). Identifying and interpreting variables. In D. R. Preston & J. G. Fought (Eds.), *American dialect research* (1st ed., pp. 193–221). J. Benjamins Pub. Co.

Zhou, J., & Xia, Y. (2024). Register-based distribution of expressions of modality in COCA. *Humanities & Social Sciences Communications*, 11(1), 613–650.

<https://doi.org/10.1057/s41599-024-03134-x>

Appendix A: Frequency of direct and euphemistic variants of 'DIE' in Southeastern United States (Johnson, 1996, p. 272)

Prompt unknown, however across questions participants were asked to supply words they would themselves use.

	1930		1990	
Variant type	Elicited forms	Percentage of elicited forms	Elicited forms	Percentage of elicited forms
direct	died, dead, died out	49% (39)	died, dead, dead and gone, deceased	33% (34)
euphemistic	DEPARTURE, SLEEP	51% (41)	DEPARTURE, (BOUNDARY CROSSING), RELEASE/DEFEAT	52% (53)
dysphemistic		0% (0)		15% (15)

Appendix B: Frequency of *PASS*-type variants of 'DIE' in in Southeastern United States (Johnson 1996, p. 272)

Prompt unknown, however across questions participants were asked to supply words they would themselves use.

	1930	1990
Elicited <i>PASS</i> -type variant	Percentage of elicited <i>PASS</i> -type variant	
pass(ed) away	73% (27)	50% (15)
passed out	22% (8)	0% (0)
passed on	5% (2)	33% (10)
passed	0% (0)	17% (5)