Hypercorrection in English: an intervarietal corpus-based study[1]

PETER COLLINS
University of New South Wales
(Received 2 January 2021; revised 9 June 2021)

This article aims to provide a fresh approach to the study of hypercorrection, the misguided application of a real or imagined rule – typically in response to prescriptive pressure – in which the speaker's attempt to be 'correct' leads to an 'incorrect' result. Instead of more familiar sources of information on hypercorrection such as attitude elicitation studies and prescriptive commentary, insights are sought from quantitative and qualitative data extracted from the 2-billion-word Global Web-based English corpus (GloWbE; Davies 2013). Five categories are investigated: case-marked pronouns, *-ly* and non-*ly* adverbs, agreement with number-transparent nouns, (extended uses of) irrealis *were*, and 'hyperforeign' noun suffixation. The nature and extent of hypercorrection in these categories, across the twenty English varieties represented in GloWbE, are investigated and discussed. Findings include a tendency for hypercorrection to be more common in American than in British English, and more prevalent in the 'Inner Circle' (IC) than in the 'Outer Circle' (OC) varieties (particularly with established constructions which have been the target of institutionalised prescriptive commentary over a long period of time).

Keywords: hypercorrection, corpus, prescriptive, GloWbE, World English

1   Introduction

Linguistic hypercorrection occurs when a real or imagined rule – involving a grammatical construction, word form, spelling or pronunciation – is applied inappropriately, with the consequence that 'excessive striving for correctness … leads in fact to incorrectness' (Huddleston & Pullum *et al.* 2002: 1680). In a similar vein, hypercorrection is defined by Decamp (1972: 87) as 'an incorrect analogy with a form in a prestige dialect which the speaker has imperfectly mastered', and by Quirk *et al.* (1985: 14) as occurring when '[a]s an occasional consequence of prescriptive pressures, some speakers have mistakenly extended particular prescriptive rules in an attempt to avoid mistakes'.

Some examples of hypercorrection follow representing the five categories investigated in the present study (see section 4), all from GloWbE (Davies 2013; the corpus that is our primary source of quantitative information: see section 3).[2]

(1) I don't think this difference between you and **I** exists (AU)
(2) Viewed **thusly**, the amount of charity expressed daily is kind of remarkable (US)

---

[2] All examples are taken from GloWbE: in each case the relevant subcorpus is indicated by a country label.

(3) Ehmke walked up to Mack and asked if he **were** still going to pitch (GB)

(4) A number of plans **is** required to cope with each of the phases (NG)

(5) **Octopi** can feel and taste with their many arms (NZ)

In all cases the highlighted hypercorrect form has a counterpart that would be more usually found in standard usage: *me* in (1), *thus* in (2), *was* in (3), *are* in (4) and *octopuses* in (5).

The aim of the study is to venture into hitherto unexplored territory, transcending more familiar sources of information on hypercorrection such as attitude elicitation studies (e.g. Mittins *et al.* 1970; Lukač & Tieken-Boon van Ostade 2019) and prescriptive commentary – in guides such as Partridge's *Usage and abusage: A guide to good English* (1963), Gowers' *Fowler's dictionary of modern English usage* (1965) and Garner's *A dictionary of Modern American usage* (1998) – with a comprehensive corpus-based account which embraces the various categories of hypercorrection in English, and which explores the nature and extent of hypercorrection not just in British English (BrE) and/or American English (AmE) but in the complete set of all twenty World Englishes represented in GloWbE.

The structure of the rest of the article is as follows. Section 2 discusses the interrelationship between hypercorrection and sociolinguistics, hypercorrection and prescriptivism, and hypercorrection and second language acquisition (SLA). Section 3 presents the design of GloWbE. Section 4 presents the corpus-derived findings for the five types of hypercorrection. Section 5 is devoted to discussion and concluding remarks.

2 Hypercorrection in relation to various fields of study

2.1 Hypercorrection and sociolinguistics

Hypercorrection has been explored in studies of language variation and linguistic change. As characterised by Labov (1966, 1972), and invoked in sociolinguistics (e.g. Wolfram 1991) and in historical linguistics (e.g. Campbell 1998), hypercorrection is understood to be prompted by speakers' awareness of – and their linguistic insecurity regarding – differing degrees of prestige associated with language varieties, with the consequent production of forms mistakenly thought to match more prestigious patterns resulting in language change.

In addition to the speaker variable of social class, as explored in Labov's seminal sociolinguistic research, other variables that have been found to impinge on hypercorrection are age and educational level. For example, in Angermeyer & Singler's (2003) study of pronominal case variation, it was found that standard patterns were favoured by speakers who were older and more educated than those who favoured the hypercorrect pattern exemplified in (1) above. Unfortunately, socio-demographic information is not available for the speakers whose texts are included in GloWbE, although the design of the corpus does enable us to make assumptions about likely country of origin (and, therefore, likely native-speaker or non-native-speaker status: see

section 2.3 below). Consequently, the study's findings are more relevant to the study of World Englishes than to sociolinguistics.

2.2 Hypercorrection and prescriptivism

I have invoked speakers' attempts to apply principles associated with linguistic 'correctness' and 'prescriptivism' in defining and explaining hypercorrection. How do speakers become exposed to such principles? For many speakers the primary exposure no doubt derives from their institutionalised educational experience. According to Crystal (2006), such educational exposure has waned in recent decades, in response to the rising influence of descriptive linguistics (the practice of describing and explaining observed language phenomena without evaluating them). He describes the implementation in the 1990s of a 'pragmatic approach' in English classroom pedagogy which 'replaces the concept of "eternal vigilance" (beloved of prescriptivists and purists) by one of "eternal tolerance"' (2006: 410). Crystal's prediction was that by focusing on the variability of linguistic rules in context, teachers would be able to eliminate prescriptive tenets and their social consequences, thereby helping young Anglophones become tolerant of language variation and change. There is, however, some empirical evidence which indicates that Crystal's prediction has not come true (see Burridge 2010; Severin 2017).

In addition to the 'top-down', institutional, promotion of prescriptivism there is a tradition of 'bottom-up', public, attempts to address putative misuses of language and linguistic decline, referred to as the 'complaint tradition' by Milroy & Milroy (2012), and as 'grassroots prescriptivism' by Lukač (2018a; 2018b). Lukač argues vigorously – like Cameron (1995), Hundt (2009) and Davies & Ziegler (2015: 4) before her – that theoretical models of language standardisation in linguistics have tended to underestimate the contribution of the grassroots prescriptivism as manifested traditionally in letters to newspapers, public forums and, more recently, social media (to which we may add grammar checkers in software programs, as discussed by Curzan (2014: 64–92)).

2.3 Hypercorrection and SLA

In the Second Language Acquisition (SLA) literature hypercorrection is occasionally considered to be a source of errors, along with overgeneralisation, faulty teaching and fossilisation (see, for example, Touchie 1986; Eckman et al. 2013). Some consider that hypercorrection in SLA results from crosslinguistic influence, manifested in an overreaction to L1 influences (Odlin 1989: 38), a claim challenged by Eckman et al., who argue that hypercorrection in SLA derives from a perceived gap between alternative realisations of a variable (one prestigious and one non-prestigious) within the L2, not between alternative realisations of a variable in LI and L2. Eckmann et al. propose that hypercorrection represents a late stage in SLA, a suggestion that derives plausibility from studies that have found an association between hypercorrection and

educational level (e.g. Angermeyer & Singler 2003). Eckmann *et al.* also accept the prevailing view in both sociolinguistic and SLA studies that hypercorrection results from linguistic insecurity on the part of the speaker, arguably generated by speakers' confusion over divergences between local usages and older norms of correctness (see Schneider 2007: 43). This prevailing view is potentially relevant to our explanations for divergent frequencies between the L1 and L2 varieties represented in GloWbE (see further next section).

3   The GloWbE corpus

The data for the present study were extracted from GloWbE, which comprises nearly 2 billion (1,885,632,973) words of text (both 'general' texts from newspapers, magazines, company websites and the like; and blogs) collected from the web pages of twenty different countries (Davies & Fuchs 2015; www.english-corpora.org/glowbe/). For details of GloWbE's design, see table 1 (which also contains information relevant to section 5). The size of GloWbE enabled retrieval of a sufficient number of tokens to allow for the inclusion of the sometimes low-frequency items that are relevant to this study, something that would not have been possible with smaller-sized corpora, such as the l-million-word components of the *International Corpus of English* (ICE) collection (http://ice-corpora.net/ice/index.html). It is the availability of a corpus of GloWbE's dimensions that makes possible the novel application in this study of an empirically based approach to hypercorrection: relative lowness of frequency. In other words, in order for a usage to qualify as hypercorrect I shall require that its frequency in GloWbE be smaller – and preferably markedly smaller – than that of its corresponding sanctioned standard variant. The web-based texts of GloWbE, approximately half of which are blogs (q.v. Loureiro-Porto 2017: 455), are suitable for studying hypercorrection, these being precisely the type of texts in which speakers will tend to not monitor their language closely, and in which usages about which they are unsure or insecure can therefore be predicted to occur.

Table 1 presents the (GloWbE labels for the) twenty countries, and their associated English variety labels, along with the number of words in the twenty subcorpora, and subclassifications relevant to the study (explained below). I shall use the country labels when referring to the twenty GloWbE subcorpora, and the variety labels when referring in general to the Englishes represented by the subcorpora.

The subclassification labels 'Inner Circle' (IC) and 'Outer Circle' (OC) are taken from Kachru (1985), in whose 'concentric circles' model of World Englishes (WEs) the structural properties of the first or 'native' English varieties of the IC countries – the United States, Canada, Great Britain, Ireland, Australia and New Zealand – differ from those of the institutionalised second-language varieties of the OC countries – India, Sri Lanka, Pakistan, Bangladesh, Singapore, Malaysia, Philippines, Hong Kong, South Africa, Nigeria, Ghana, Kenya, Tanzania and Jamaica – as a result of such factors as language contact in L2 acquisition, differences of norm orientation, and substrate influence. The fourteen OC countries can be subdivided into four geographical

Table 1. *Classification and word count of the twenty regional varieties in GloWbE*

WEs/Kachru classification		Country	Country label	Variety label	Word count*		
					Variety total	OC zone total	IC/OC total
Inner Circle		USA	US	AmE	386,809,355		1,239,817,686
		Canada	CA	CanE	134,765,381		
		Great Britain	GB	BrE	387,615,074		
		Ireland	IE	IrE	101,029,231		
		Australia	AU	AusE	148,208,169		
		New Zealand	NZ	NZE	81,390,476		
Outer Circle	South Asia	India	IN	IndE	96,430,888	234,039,410	645,815,287
		Sri Lanka	LK	SLE	46,583,115		
		Pakistan	PK	PakE	51,367,152		
		Bangladesh	BD	BDE	39,658,255		
	South-East Asia	Singapore	SG	SingE	42,974,705	169,095,257	
		Malaysia	MY	MalE	42,420,168		
		The Philippines	PH	PhilE	43,250,093		
		Hong Kong	HK	HKE	40,450,291		
	Africa	South Africa	ZA	SthAfrE	45,364,498	203,016,954	
		Nigeria	NG	NigE	42,646,098		
		Ghana	GH	GhanE	38,768,231		
		Kenya	KE	KenE	41,069,085		
		Tanzania	TZ	TanzE	35,169,042		
	Caribbean	Jamaica	JM	JamE	39,663,666	39,663,666	
TOTAL							1,885,632,973

* Frequencies are as in: www.english-corpora.org/glowbe_corpus.asp

'zones': South Asia (SA), South-East Asia (SEA), Africa (Afr) and the Caribbean (Carib). For more discussion of the twenty GloWbE varieties, and subdivisions thereof, see Collins (2020).

Finally, some brief comments on the twenty Englishes represented in GloWbE will be offered. Within the IC, BrE and AmE are recognised as 'reference' varieties. The influence of BrE is in evidence in its role as a colonial 'parent' in the evolution of postcolonial varieties, and the influence of AmE is in evidence latterly in its strong impact on English worldwide, a reflection of the international influence of the USA. IrE and CanE have features that reflect British and American influence respectively. AusE and NZE are established Southern Hemisphere varieties with closely related histories. Of the four South Asian Englishes IndE is the most internationally well known, and institutionalised to a higher degree than its neighbours, SLE, PakE and BDE. Within South-East Asia, SingE has evolved further (see Schneider 2007: 153–61) than MalE, PhilE (the only GloWbE variety with AmE rather than BrE as colonial 'parent') and HKE. In Africa, English has to compete with a large number of local languages, in South Africa (SthAfrE), in West Africa (NigE and GhanE) and in East Africa (KenE and TanE). In the Caribbean, JamE – distinguishable from Jamaican Creole – has moved away from BrE norms since the 1960s.

4 The findings

In this article I explore the five categories in which hypercorrect variants occur, as identified in section 1 above. The selection of these categories was based on the extent of their discussion variously in reference grammars, dictionaries and usage manuals, on websites located via Google and on corpus searches applied systematically to GloWbE, and occasionally to two American diachronic corpora: the *Corpus of Historical American English* (COHA; Davies 2010–) and the *Corpus of Contemporary American English* (COCA; Davies 2008–). Salient quantitative findings will be supplied wherever possible. In the case of some variables the number of hypercorrect forms is insufficient to venture generalisations any finer than those involving a comparison of IC vs OC frequencies.

4.1 Nominative and accusative pronouns

Because prescriptive grammarians have traditionally tended to accept only formal style as 'grammatically correct', the use of accusative pronouns in certain constructions where they are associated with informal style has attracted criticism, and this has given rise to the hypercorrect use of nominative pronouns. Following Huddleston & Pullum *et al.* (2002: 458–67), I shall distinguish three subcategories. The first two involve personal pronouns (and determinatives) in, respectively, non-coordinative and coordinative constructions, while the third involves interrogative and relative *who/whom* where, by contrast with the first two subcategories, it is the accusative form that is associated with formal style.

4.1.1 Non-coordinative personal pronouns/determinatives

There are three constructions in which the use of a nominative pronoun or determinative can be regarded as involving hypercorrection, namely those in which a nominative pronoun is subject of a *for*-infinitival clause, those in which determinative *we* is selected in a non-subject NP and those in which a nominative pronoun is complement of comparative *than* or *as*.

(i) Huddleston & Pullum *et al.* (2002: 461) regard the use of nominative pronouns as subject of a *for*-infinitival clause as ungrammatical. However, there are 58 tokens in GloWbE – as exemplified in (6) – with the OC evidencing a stronger predilection for nominative forms than the IC (0.43 vs 0.29 tokens pmw).

(6) For **we** to win a bullion we have to be so self-motivated and so self-disciplined. (GB)

(ii) The hypercorrect use of determinative *we* in non-subject NPs, which is not mentioned in the major reference grammars of English, is most likely reinforced by the stigma attached to the vernacular use of the accusative pronoun *us* in subject NPs, as in (7):

(7) my mom would threaten to put his tape in the stereo when **us** kids were acting up (CA)

Direct object NP examples, as in (8), are rare in GloWbE. However, those where the NP is complement of a preposition are not uncommon, with *with* – as in (9) – displaying the highest number of tokens (109) and, as with other prepositions, the OC showing a stronger appetite for hypercorrection (0.97 tokens pmw) than the IC (0.37).

(8) our policy makers are NOT protecting **we** ratepayers. (AU)
(9) As it's the case with **we** humans, so it is with businesses too (NG)

(iii) The third non-coordinative construction in which the use of nominative forms arguably represents a hypercorrection is that where such forms serve as complements of comparative *than* as in (10), or *as* as in (11):

(10) So did his brother and sister, who were much older than **he**. (HK)
(11) If your response to terror is to feed and clothe and nurture your terrorist offspring then you are as guilty as **they**. (US)

In traditional grammars and conservative usage handbooks, *than* and *as* as used here are generally regarded as conjunctions introducing an elliptical clause, following arguments put forward as far back as the eighteenth century by Robert Lowth (1762), whose analysis of *than I* involves *am* 'being understood' after the pronoun (p.166). On this interpretation the case of the pronoun 'should' be nominative ('who were much older than he/*him was'; 'you are as guilty as they/*them are'). However, there is an alternative interpretation according to which *than* and *as* are understood to be prepositions, whose complement is simply an NP – and standardly, if pronominal, in the accusative case – rather than a reduced clause. The latter is the position adopted in authoritative reference grammars such as Quirk *et al.* (1985: 661), and in less

prescriptive usage guides, such as *Merriam-Webster's dictionary of English usage* (1994), which cautions reassuringly:

> Some people think they're better than you because they say 'better than I' instead of 'better than me.' They're not, of course. They're just among the select group of grammar enthusiasts who think that *than* can only be a conjunction. You, on the [other] hand, recognize that it can also be a preposition. (see www.merriam-webster.com/words-at-play/than-what-follows-it-and-why)

Prepositional uses have a long ancestry, receiving recognition from grammarians as early as the eighteenth century (q.v. the 'topical glossary' in Leonard 1929). In line with the prepositional analysis, the nominative pronoun after *than* and *as* is here considered to be a hypercorrect variant of the corresponding accusative pronoun, as it is with other prepositions. This interpretation is supported by the trend for the accusative variant to be far more common than the nominative variant (by a ratio in GloWbE of 12.07:1 for *than*, and 9.15:1 for *as*). On the reduced clause interpretation the more rarely occurring nominative form would have to be regarded, counterintuitively, as the established standard variant, with the accusative alternant presumably belonging to a colloquial dialectal variety.

GloWbE searches were conducted for 'than P.' and 'as P.', the full stop being included to block tokens where *than* and *as* are followed by a clause as in *than I did* where only the nominative forms are possible: the relevant manifestations of P were accusative *me*, *us*, *him*, *her* and *them*; and nominative *I*, *we*, *he*, *she* and *they*. On two measures hypercorrection was found to be more common in the IC than in the OC: average pmw frequencies for the nominative forms (*than*: IC 5.54 vs OC 2.09; *as*: IC 3.06 vs OC 0.98); and accusative vs nominative ratios (*than*: IC 11.39:1 vs OC 13.54:1; *as*: IC 8.72:1 vs OC 10.52:1). One factor that is likely to play a role in these differences is the persistence and pervasiveness of prescriptivism in the IC, and the consequently widespread confusion over, and misunderstanding of, the recommendations disseminated by prescriptive 'authorities' (see, for example, Drake 1977; Pullum 2009): see section 5 for further discussion. A similar explanation can plausibly be applied to the status of AmE as the variety most prone to hypercorrection in the IC (and, in turn, to the strong nominative frequencies in PhilE, in light of continuing American influence in the Philippines). A search of COHA and COCA revealed two complementary trends in recent decades: a decline in nominative frequencies and an increase in accusatives (which may, in light of the popularity they enjoy in the TV and movie genres of COCA, be plausibly linked to the colloquialisation of Modern English: see, for example, Mair 2006 and Collins & Yao 2019).

4.1.2 *Coordinative constructions*

There is a stigma associated with accusatives in subject coordinations such as *John and me will help you*, and even more so (because of the 'impolite' prominence given to the speaker) in those with *me* as first coordinate, as in *Me and John will help you*. People who are taught that such subject coordinations are incorrect are likely to 'generalise their avoidance of such coordinate accusatives to other functional positions' (Huddleston & Pullum *et al.* 2002: 463), thereby producing sentences such as:

(12) For some reason I can see Ryan and **I** as parents of a baby (US)

(13) Do the banks give it to you or **I**? (CA)

(14) During homework time is a great opportunity for **he** and I to collaboratively organize the binder. (US)

The status of nominative forms in examples such as (12) as hypercorrections – argued for below on frequency grounds – is recognised by Quirk *et al.* (1985: 338). Huddleston & Pullum *et al.* (2002: 463), however, contend that examples like (12), with a nominative as final coordinate in object position, do not involve hypercorrection, being 'so common in speech and used by so broad a range of speakers that it has to be recognised as a variety of Standard English'. They do nevertheless allow as hypercorrection cases where a nominative occurs in its initial position (as in 'I can see *I* and Ryan'), where it is far less frequent. The use of nominative pronouns as the complement of a preposition, as in (13), here regarded as hypercorrect, has regularly attracted prescriptive censure, as in (Fowler & Fowler 1906: 73) where it is described as 'a bad blunder'. In the so-called 'accusative and infinitive' construction in (14), the subordinator *for* shares a number of properties with its historical source, the preposition *for* (see Huddleston & Pullum *et al.* 2002: 1181–3). Accordingly, in interpreting the use of nominative forms here as involving hypercorrection, the same criteria can be applied that were invoked with respect to (6) above.

There follows a presentation of the findings for the three categories of coordinative constructions distinguished above.

(i) Coordination as object of a verb

The six tokens of this construction in GloWbE, including (12) above, and (15) and (16) below, are so rare as to invalidate intervarietal comparisons. In some cases, including (16), the dysfluent syntax suggests that L2 errors may be involved.

(15) She introduced **he** and Jeremy. (US)

(16) Chen Fang waited a person to early tell **he** or **she** (NG)

(ii) Coordination as complement of a preposition

The only preposition in GloWbE that yields a number of coordinated complements – and more particularly a complement with one or two nominative pronouns as complement – that would be sufficient to sustain a hypercorrection analysis, is *between*. Let us consider the most frequent and widely discussed type of *between*-phrase, that with *you* as first coordinate and *I* – instead of *me* – as second coordinate, as in (17).[3]

(17) Seriously, between you and **I**, there are just some experiences that makes us as Jamaicans truley unique. (JM)

---

[3] Further confirmation of the popularity of this type is provided by Quinn's (2005) elicitation study, which found greater tolerance from her respondents towards positional flexibility with first-person singular pronouns than with other person/number pronouns.

The view expressed by Huddleston & Pullum *et al.* (2002: 463) and others that examples of this type do not involve hypercorrection, referred to above, is here rejected on the basis of frequency ratios: *me* dominates over *I* in both the IC (by a ratio of 4.14:1) and in the OC (6.05:1). The notable difference between the IC and OC here is reflected as well in pmw frequency differences, the IC yielding three times more tokens of *between you and me* (1.11 pmw) than the OC (0.37 pmw). Within the IC, the strongest preference for the accusative variant occurred in AmE, the weakest in BrE.

(iii) Coordination as subject of a *for*-infinitival clause

This construction is different from the non-coordinative accusative and infinitive construction discussed in section 4.1.1 (i) above. In this case there is arguably more pressure to select a nominative form as the second coordinate, because it is further from subordinator *for* (which 'predisposes' accusative selection) and closer to the verb (which 'predisposes' selection of nominative case for the clause subject). Across all twenty varieties, excluding pronouns that are not case-marked (e.g. *myself*, *someone*, *mine*), and regardless of whether there is a valid pronoun in just one or both of the coordinates, the numbers of tokens in GloWbE were 34 nominatives vs 100 accusatives overall (5 nominatives vs 59 accusatives in the first coordinate, and 29 nominatives vs 41 accusatives in the second coordinate). The striking contrast here is the relative unpopularity of nominatives in the first coordinate, vis-à-vis the second coordinate. Out of the four possible combinations, the number of occurrences of each were: nominative + nominative (5 tokens), as in (18); accusative + accusative (25 tokens), as in (19); accusative + nominative (11 tokens), as in (20); and nominative + accusative (no tokens). The conclusion that can be drawn from these numbers is that accusatives are preferred more in the first than the second coordinate, and nominatives are preferred more in the second coordinate than the first.

(18) it would be too easy for **she and I** to swap info (GB)
(19) Lucas tries to distract her and offers for **her and him** to do a fund raiser (US)
(20) If anything was destined, it was for **him and I** to be together (PH)

There was a stark contrast between the IC and the OC frequencies, with the former accounting for 15 of the 16 tokens of hypercorrect nominative pronouns in the permissible combinations (the only OC token being (20)).

4.1.3 Who *and* whom

Despite the insistence of prescriptivists that nominative *who* should be used in the subject function only, and that accusative *whom* is the only 'correct' form in functions other than the subject, uses of *whom* now deemed 'incorrect' have a long history – as observed by Aarts (1994: 71) – as in Shakespeare's *The Tempest*, where Prospero's 'And in these fits I leave them, while I visit Young Ferdinand, whom they suppose is drown'd' has *whom* functioning as subject of the embedded clause 'X is drown'd'.

Both relative and interrogative *whom* are subject to hypercorrect use, as in (21) and (22) respectively:

(21) He cited the experience of his wife, **whom** he said became pregnant despite taking contraceptives. (PH)

(22) **Whom** do you think will win this race to provide Kenyans with affordable mortgages and by extension housing? (KE)

In both examples *whom* functions as the subject of an embedded content clause: 'X became pregnant …' and 'X will win this race …'. Such examples suggest that there are speakers who will, when confronted with syntactically complex constructions involving embedded clauses, anticipate that there might be something wrong with *who* (a misapprehension prompted by their vague awareness of traditional proscriptions of the use of *who* in sentences such as *Who did you see?*), and consequently select hypercorrect *whom*. The case for hypercorrection is stronger for interrogative *whom* than for relative *whom*. In the latter case, in fact, contrary positions are adopted in the two most authoritative grammars of contemporary English. Huddleston & Pullum *et al.* (2002: 466–7) do not accept this use of *whom* to be hypercorrect, arguing that:

> The accusative variant has a long history and is used by a wide range of speakers; examples are quite often encountered in quality newspapers and works by respected authors. It has to be accepted as an established variant of the standard language. Thus there are in effect two dialects with respect to the case of embedded subjects, though they are not distinguished on any regional basis.[4]

The position adopted by Quirk *et al.* (1985: 368) – and in the present study, on quantitative grounds: see below – is that the relative use of *whom* discussed above does in fact involve hypercorrection.

Relative *who(m)* frequencies yielded by the routine 'who(m) he SAY V' (in prenuclear position preceding the subject of a relative clause) indicated a preference for *who* over *whom* of 446:163, or 2.73:1. The mildness of this ratio might be interpreted by some as confirming that this use of *whom* is sufficiently established in the language to invalidate a hypercorrection interpretation. However, I would argue that such a conclusion is undermined by the 0.86 pmw GloWbE frequency for *whom* in this construction, one so miniscule that the variant with *whom* cannot be regarded as 'established', and that a hypercorrection interpretation is therefore defensible. A comparison of the IC and the OC varieties indicates that the former (whose *who* vs *whom* ratio was 3.7:1) is less prone to hypercorrection than the latter (whose ratio was 2.1:1). It may be suggested that the differences are ascribable to greater uncertainty over the interpretation of the syntax amongst L2 speakers of English in the OC, than L1 speakers in the IC.

A rather less controversial type of hypercorrection with relative *whom*, one accepted as hypercorrection also by Huddleston & Pullum *et al.* (2002: 467n), occurs when *whom* functions as the subject not of an embedded clause, but rather of the relative clause

---

[4] Huddleston & Pullum *et al.*'s position is rendered questionable by a search of COCA for 'who(m) he says/said V' which found the *whom* variant to be barely 'established' in newspapers, where it is outnumbered by a ratio of 5.1:1, and in fiction by 10.0:1.

itself, with a parenthetical expression after the subject, as in (23). A search for 'who(m), P BELIEVE,' yielded a mere ten instances for *whom* as against 88 for *who*, and for 'who (m), P THINK,' yielded only two for *whom* as against 68 for *who*.

(23) And that's a question that's particularly pertinent to expats, **whom**, I believe, almost permanently suffer some degree of FOMO. (GB)

GloWbE search results for the routine 'who(m) do you think V' support the claim made above that the case for hypercorrection is more compelling with interrogative than relative *whom*. The overall *who* vs *whom* ratio was 1502:24 (or 62.5:1), the 24 tokens (0.012 pmw) of *whom* contradicting Huddleston & Pullum *et al.*'s (2002: 466) observation that '[t]he accusative construction … does not appear to occur in main clause interrogatives'. As in the case of relative *whom*, so with interrogative *whom*, the OC varieties are considerably more prone to hypercorrection than the IC, as measured both by the frequency of this construction (OC 0.020 pmw vs IC 0.008 pmw), and by *who* vs *whom* ratios (IC 94.5:1 vs OC 35.5:1).

A handful of instances of *whom* in subordinate interrogatives were observed in GloWbE, including (24), a use described by Huddleston & Pullum *et al.* (2002: 466) as 'rare and of doubtful acceptability' (their example being [?]*I told her whom you think took it*).

(24) Terry had turned to me at halftime and asked **whom** I thought would win. (US)

Finally, hypercorrect tokens of non-referential *whomever* (for *whoever*, in subject function), as in (25), were observed in GloWbE, a use not noted in comprehensive English grammars. The application of a search routine that would capture relevant tokens ('. whomever V'), albeit only in sentence-initial position, yielded 86 tokens, a frequency outstripped by that for *whoever* by a ratio of 52:1.

(25) **Whomever** is the next president should therefor [sic] be encouraged to immortalize themselves by leading this effort. (US)

In this construction, hypercorrect *whomever* is far more popular in the IC (0.057 tokens pmw) than in the OC (0.023). It is remarkably popular in US (0.108), possibly reflecting the strong rejection in grassroots American prescriptivism of *who* – and by extension *whoever* – in non-subject functions (see, for example, www.grammarbook.com/blog/pronouns/whom-abuse-is-rampant/).

4.2   *-ly* and non-*ly* adverbs

A large number of English adverbs are derived from adjectives via *-ly* suffixation, a fact that has given rise to the mistaken assumption that all adverbs should end in *-ly*, with the resultant hypercorrect selection of *-ly* adverbs when such variables exist. Non-*ly* adverbs that are homonymous with adjectives (e.g. *fast*, *hard*, *slow*; but not *often*, *ever*, *perhaps*, *thus*, *moreover*, etc.) are often referred to as 'flat' adverbs. Flat adverbs were relatively common in the eighteenth century, but their frequency subsequently declined due to

the stigma deriving from their censure by grammarians who believed them to be adjectives and insisted that they be replaced by their *-ly* counterparts (see further Nevalainen 1997). According to frequencies presented by Biber *et al.* (1999: 540), *-ly* adverbs are about 1.3 times more common than 'simple' adverbs. The distaste for flat adverbs continues in the modern era, with Biber *et al.* observing that '[f]rom a prescriptive point of view, this use of the adjective form is often stigmatized as non-standard' (1999: 542) and Huddleston & Pullum *et al.* (2002: 567) similarly referring to 'recognisably non-standard uses'.

Undoubtedly major factors in the variable usage of *-ly* adverbs are speaker misunderstanding and uncertainty. The first of these is characterised by Butterfield (2015: 1131–2) as follows: 'Whenever a single-syllable adjective is used in an obviously adverbial role some people suffer from what might be called the "absent *-ly*" or "something-is-missing syndrome". Because so many single-syllable adjectives (*apt*, *brief*, *damp*, etc.) are never used as adverbs, it is an easy step to believing that none can be used.' The latter factor is invoked by Quirk *et al.* (1985: 407), who observe the existence of 'uncertainty in the use of adjective and related adverb forms' in cases such as *The flowers smell sweet/?sweetly* and *He felt bad/?badly about it*.

For the purposes of the present study I have identified three subcategories of *-ly* adverbs, distinguishable via quantitative and qualitative contrasts with their non-*ly* counterparts: (i) those with the same meaning as their non-*ly* counterparts, but comparatively rare; (ii) those with the same meaning as their non-*ly* counterparts, but comparatively more established in the language than those in (i); and (iii) those with limited overlap of meanings/uses with their non-*ly* counterparts. In all three subcategories the use of the *-ly* adverbs is argued to be triggered, in varying degrees, by hypercorrection. In the ensuing discussion, for each of the three categories in turn, I shall begin by listing the adverbs that meet the criteria and which corpus searches confirmed to have appropriate and/or viable frequencies for inclusion, and then proceed to present examples, quantitative findings and discussion.

4.2.1 Type 1 -ly adverbs

Leastly (10), *longly* (19), *nextly* (12), *otherwisely* (3), *soonly* (14), *welly* (20), *worsely* (8)

The seven items in this category are very uncommon, not only in terms of raw frequency (in the present study a maximum of 20 tokens in GloWbE was set: frequencies for each item are indicated in brackets), but also vis-à-vis their non-*ly* counterparts. Accordingly, there is a strong case for ascribing their occurrence to hypercorrection, although in some cases they might alternatively be regarded as errors or spurious neologisms, especially in L2 OC varieties. The seven selected items are all rare in contemporary English: *longly*, *nextly*, *soonly* and *welly* all being given an *Oxford English Dictionary* (*OED*) Band 2 rating – applicable to words whose frequency is lower than 0.01pmw in typical modern English usage, and which 'are not part of normal discourse and would be unknown to most people': see https://public.oed.com/how-to-use-the-oed-key-to-frequency/ – while *leastly*, *otherwisely* and *worsely* do not receive entries. Furthermore,

the four adverbs with *OED* entries all have long ancestries, with citations dating back to the fifteenth century for *longly* and *soonly*, to the sixteenth century for *nextly* and to the seventeenth century for *welly*.

There is a formally distinct subclass of Type 1 adverbs ending in -*lily*, often mentioned in usage guides (see Tieken-Boon van Ostade 2020: 142–4), that were excluded from the present study on the grounds of their extreme rareness. Some are recognised in the *OED*, and of these *friendlily* has the most tokens in GloWbE (6), followed by *sillily* (4), *uglily* (4) and *lonelily* (1). As Quirk *et al.* (1985: 1556) observe, 'more usually, prepositional phrases or synonyms are used', such as *in a friendly way* or *amicably* instead of *friendlily*.

For the category of Type 1 -*ly* adverbs, the paucity of the numbers involved requires us to exercise caution in making intervarietal comparisons based on anything more than the broadest groupings (IC vs OC), except in the case of the category as a whole: see table 2. It may also be noted that the search results for two items were inflated by a large proportion of irrelevant tokens, which had to be manually eliminated: *worsely* as the proper name *Worsely*; and *welly* as a diminutive noun referring to the New Zealand city *Wellington*, as a colloquial abbreviation of *Wellington* (*boot*) and as a verb meaning 'to give something the boot'. Adverbial *welly* is commonly used facetiously in NZE, unsurprisingly in view of the popularity of puns and banter relating to the nouns *Welly* and *welly* in New Zealand: see e.g. www.cartoonstock.com/directory/w/wellington_boots.asp. Some Type 1 examples follow:

(26) I had **longly** touted the Lakers' need to keep Bynum (US)
(27) Maybe in a legal market (which hopefully will be widespread **soonly**) the best idea is to grow ganja in an appropriate sized greenhouse (CA)
(28) The exit is the hard part which you know and have dealt with very **welly**. (NZ)

The OC proved to be more prone to hypercorrection than the IC, by a ratio of 1.39:1. Within the IC, BrE led the way, while in the OC South-East Asia and the Caribbean were relatively averse to hypercorrection, with frequencies even smaller than that for the IC.

4.2.2   Type 2 -ly adverbs

Fastly (131), *muchly* (125), *oftenly* (744), *outrightly* (244), *seldomly* (91), *straightly* (111), *thusly* (740), *uprightly* (95)

The -*ly* adverbs in this category yielded GloWbE frequencies which, though larger than those of Type 1 -*ly* adverbs, are considerably less than 50 per cent of those of the corresponding non-*ly* adverb. While the use of Type 1 -*ly* adverbs is certainly not established in Standard English, that of Type 2 -*ly* adverbs is a matter of contention. It could be argued that the use of Type 2 -*ly* adverbs is sufficiently established in English for them to be regarded as merely (non-hypercorrect) dialectal variants in the language. However, the hypercorrection-based position adopted here derives support both from the frequencies presented below and from the number of prescriptive criticisms that have targeted Type 2 -*ly* adverbs. For example, in grassroots commentary *oftenly* has

Table 2. *Pmw frequencies for Type 1 -ly adverbs in GloWbE*

US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK	ZA	NG	GH	KE	TZ	JM
0.04	0.05	0.06	0.02	0.04	0.04	0.07	0.04	0.04	0.21	0.04	0.00	0.02	0.05	0.10	0.06	0.03	0.00	0.12	
							0.090				.027					0.062			0.03
			0.041									0.057							

been denounced as an illegitimate formation (www.englishforums.com/English/OftenAndOftenly/vqqkm/post.htm) and *fastly* labelled a 'blooper' (www.rediff.com/getahead/2007/jun/08eng.htm). In usage guides *thusly* – which has 26 entries in the HUGE database (for information of the nature and use of HUGE see Straaijer 2014) – has been pilloried for being a 'nonword' and its use 'a serious lapse' (Garner 1998: 654), as 'unnecessary' (Allen 1999: 573), as an 'abomination' **(**Morris & Morris 1975: 599) and as typical of the 'poorly educated' (Pickett *et al.* 2005: 464). For further discussion of the – characteristically censorious – treatment of *thusly* in usage guides, see Lukač & Tieken-Boon van Ostade (2019: 174–5) and Lukač (2018c).

Some Type 2 *-ly* adverbs occur predominantly in particular contextual varieties (or 'registers') and regional varieties (or 'dialects'). The most clearcut instance of the former is *uprightly*, which scrutiny of the 95 tokens in GloWbE indicated is used mainly to modify the verbs *walk* and *act* in theological and scriptural contexts where the *-ly* suffix arguably adds an archaic biblical dimension, as in (33) below. Accordingly, it was unsurprising that GloWbE frequencies were dominated by US counts in the IC and by NG in the OC (the USA has the largest Christian population in the world and Nigeria the fifth largest: see https://en.wikipedia.org/wiki/List_of_religious_populations#Christians). *Muchly* and *thusly* are commonly found in lighthearted and humorous situations: the former – said to be 'used humorously' by Allen (1999: 411) – as in (30); the latter well-known as a result of its use in Sheldon's *I have informed you thusly* in the American sitcom *The Big Bang Theory*. This distribution is potentially relevant as an explanation for the relative infrequency of their occurrence in the OC varieties of GloWbE, where there may be limited awareness of the jocular use of these adverbs.

Thusly, the most famous – or perhaps infamous – hypercorrect *-ly* adverb, enjoys far greater acceptance in AmE than in BrE, the present study finding *thusly* to be almost three and a half times more frequent (3.42:1) in US than in GB (leaving us with a case for hypercorrection in AmE that is at best tenuous). Corroboration of the AmE vs BrE difference is available in the findings of Lukač & Tieken-Boon van Ostade's (2019) attitude study and Butterfield's (2015: 818) impressionistic observation regarding the 'bemused derision with which many BrE speakers are likely to greet it'. In other cases a British penchant is in evidence: for example, the GloWbE-GB frequency of *muchly* (0.13 tokens pmw) exceeds the IC average of 0.08, and *oftenly* is more popular in GB than in other varieties.

GloWbE frequencies for the eight Type 2 *-ly* adverbs are presented in table 3. Some examples follow:

(29) Yes, she reply but her heart pumped **fastly**. (PK)
(30) thank you for your lovely comments which make me smile **muchly** in the month of mad mad madnesses (GB)
(31) He may not be lying **outrightly** but he doesn't care to be truthful. (NG)
(32) He summed up the event **thusly**: 'I was just trying to have some fun.' (US)
(33) It is needful, in the first place, to act **uprightly** in the sight of God (US)

Table 3. *Pmw frequencies for Type 2* -ly *adverbs in GloWbE*

US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK	ZA	NG	GH	KE	TZ	JM
1.25	0.80	0.57	0.50	0.53	0.67	0.99	0.90	1.17	0.67	0.91	0.59	1.14	0.49	0.45	2.40	0.57	1.01	0.87	
							0.93				0.78					1.06			0.69
		0.72										0.91							

The OC is again more prone to hypercorrectness than the IC (by a ratio of 1.26:1), with OC numbers boosted by the exceptionally high frequency for NG (particularly for *outrightly* and *uprightly*). The IC average in turn is boosted by the very high score for *thusly* in US and the high score in CA.

4.2.3   Type 3 -ly adverbs

badly, cheaply, cleanly, clearly, closely, dearly, deeply, fairly, hardly, highly, lightly, loudly, lowly, rampantly, rightly, safely, slowly, sweetly, tightly, wrongly

This is the largest subcategory in the study, reflecting the prevalence of cases in which hypercorrection does not apply across the full spectrum of collocations for a particular adverb, but rather to a small subset thereof, where the hypercorrect *-ly* form is significantly less frequent than its non-*ly* counterpart. For example, the preference for the adverb *safely* rather than *safe* is clearly not motivated by hypercorrectness when it modifies the verbs *drive* and *travel*, but arguably it is so-motivated in collocation with *keep*, where the default selection is *safe*. Or again, the case for hypercorrectness as motivation for the use of *cheaply* is more plausible when it collocates with *come* (where there are far fewer tokens of *cheaply* than of *cheap*) than with *buy* (where there are almost as many tokens of *cheaply* as there are of *cheap*).

The relevant use(s) of each *-ly* adverb, for which GloWbE searches and follow-up inspection of KWIC outputs were performed to determine the most frequent hypercorrect collocates, were as follows ('+' means 'collocates with' or 'modifies'): *badly* [+ *feel*, *smell*]; *cheaply* [+ *come*]; *cleanly* [+ *out of*, *over*, *shaven*]; *clearly* [in *loudly and clearly*]; *closely* [+ *hold*, *run*, *stay*, *stick*]; *dearly* [+ *hold*]; *deeply* [+ *cut*, *delve*, *drill*, *etch*, *go*, *look*, *penetrate*, *reach*, *run*, *sink*]; *fairly* [+ *play*, *fight*, *bid*]; *hardly* [+ *work*]; *highly* [+ *rank*, *score*]; *lightly* [+ *eat*, *pack*, *travel*]; *loudly* [in *loudly and clearly*]; *lowly* [+ *rank*, *score*, *value*, *seed*]; *rampantly* [+ *run*]; *rightly* [+ *do*, *guess*, *go*, *treat* and in *rightly or wrongly*]; *safely* [+ *keep*, *stay*]; *slowly* [+ *more* and *drive*, *run*, *go*]; *sweetly* [+ *smell*, *taste*]; *tightly* [+ *shut*, *sleep*, *sit*, *stretch*, *hold*]; *wrongly* [+ *hear*, *go*, *spell*, *guess*]. Some examples follow:

(34) I feel **badly** saying that because I love Africa. (AU)

(35) Don't expect your arena to come **cheaply**. (NZ)

(36) My father had to work **hardly** in the farm (BD)

(37) Local brands rank **highly** among the pages most engaged with. (GB)

(38) Say it **loudly** and **clearly**, and eventually it will start hitting prime time television. (US)

(39) the heather […] made the light breeze taste **sweetly** (AU)

For this category I won't present a single table comparing all twenty members, because overall corpus frequencies are not relevant. Suffice it to say that for the majority of the Type 3 *-ly* adverbs, frequencies for hypercorrect uses in the IC outstripped those in the OC (with *badly, cheaply, cleanly, clearly, closely, deeply, fairly, highly, lightly, loudly, rightly* and *safely*), while the OC outstripped the IC in the case of *dearly, hardly, lowly, rampantly, slowly, sweetly, tightly* and *wrongly*. A factor that appeared to be at play in

the latter case was the insensitivity of OC speakers to the unidiomaticity of the *-ly* adverb in the collocations examined, as in (36).

In the case of some items, particularly strong frequencies were evidenced by one or other of the reference varieties, BrE (*fairly*, *highly*, *rightly* and *slowly*) and AmE (*badly* – frequencies for which were also noted to be high for varieties in which AmE influence is found, namely CanE, PhilE and JamE – and *loudly and clearly*).

4.3 Subject–verb agreement with number-transparent nouns

As observed by Huddleston & Pullum *et al.* (2002: 501–7), the rule of number agreement between a subject and verb may be overridden if semantically motivated by a collective noun (as in 'the committee *were*'), or a number-transparent noun as in (40):

(40) A number of things **are** helping to build anticipation (GB)

In such cases the override is normally assumed to be obligatory, but occasional examples in which it is not applied do nevertheless occur, as in (41).

(41) A number of its postures **is** believed to be quite effective for relieving back pain. (HK)

GloWbE searches confirmed the accuracy of Huddleston & Pullum *et al.*'s (2002: 502) claim that such cases are too rare to qualify as an established variant in Standard English, and can therefore 'be regarded as hypercorrections attributable to an overzealous application of the simple agreement rule'. The search routine used, '. a number of N is', was designed to capture only sentences with an initial number-transparent phrase, thereby avoiding irrelevant tokens such as 'The thought of having to live for a number of years is horrific' (GB). It yielded 79 tokens, considerably fewer than those with *are* (959). Frequencies of hypercorrect *is* – and *is* vs *are* ratios – were almost identical across the IC and OC varieties.

There were several further number-transparent nouns that yielded a small number of hypercorrect hits with a singular verb in GloWbE, including: *lot* (six tokens), *few* (seven tokens) and *majority* (six tokens), as exemplified below:

(42) That's the question a lot of people **is** asking lately. (US)
(43) you ought to be aware that a few of these **is** going to charge a fee (PH)
(44) the majority of people **is** too set in supporting a premiership team. (GB)

4.4 Extended uses of irrealis *were*

In this section I shall explore the use of what Quirk *et al.* (1985: 158n) refer to as 'pseudo-subjunctive *were*', in other words irrealis *were* as used not in a modal remoteness construction of the type with which it is conventionally associated – as exemplified in (45) – but rather instead of *was* in various backshift and past time constructions where there is no suggestion of counterfactuality, as in (46), (47) and (48).

(45) If he **were** around today I would seat my family in the row either in front of him or behind and give it our all. (NZ)

(46) she even asked if it **were** possible for her to never have sex with Jim. (GB)

(47) Women were property, not people. If she **were** raped, she was either killed or forced to marry her rapist. (US)

(48) She shifted her position, as if she **were** about to go. (AU)

This practice is regarded by Huddleston & Pullum *et al.* (2002: 87) as an extension to 'certain neighbouring constructions' of speaker antipathy towards – and prescriptive censure of – the use of *was* in modal remoteness constructions. However, *was* would be the expected standard form in subordinate interrogatives as in (46), where *if* is an interrogative rather than conditional subordinator and where it could be substituted by *whether*. In this case, a further factor prompting the hypercorrect selection of *were* might well be speaker uncertainty generated by the two different but related functions of *if* (reflective of the semantic connection between interrogativity and conditionality that is grounded in mutual non-affirmativity: see further Huddleston & Pullum *et al.* 2002: 971–2). Corbett (2008), who proscribes the remoteness use of *was* and prescribes its interrogative use, refers facetiously to speakers' confusion between these uses as 'subjunctivitis'. The use of *were* in subordinate interrogatives is generally treated as incorrect or archaic in usage manuals (such as Partridge 1963: 361).

The rather rare construction containing *were* that is exemplified in (47) is an open rather than remote conditional. The corresponding remote conditional requiring *would* in the apodosis, *If she were raped, she would be either killed or forced to marry her rapist*, conveys a necessary suggestion of counterfactuality, one that is not found in the original example, where there is no suggestion that 'she' wasn't raped. In fact, the *if* in (47) is more similar semantically to *when(ever)* than it is to conditional *if*.

In comparative constructions with *as if* (and *as though*) as in (48), the use of irrealis *were* once more seems to be semantically unmotivated (there is no suggestion of counterfactuality, no suggestion that 'she' wasn't about to go). Once again, then, *were* is simply a more formal variant of *was*, one that is felt by some speakers to be more appropriate given the general association between *were* and *if* in its prototypical remote conditional use.

I shall regard the use of *were* in (46) and (47), but not in (48), as involving hypercorrection. This interpretation is based on the finding that the frequency of *were* in (46) and (47) is significantly lower than that of *was*, but in (48) they are too similar for one variant to be regarded as the dominant standard and the other as a less readily accepted hypercorrection (*as if … were* 2023 vs *as if … was* 1947).

A search was conducted for relevant embedded interrogatives with the routine 'asked if P was/were' (*ask* being more common than any other verb of inquiry), and the results manually searched to eliminate plural pronouns, ultimately being limited to instances with *he*, *it*, *I*, *she* and *anything* as subject. The *was* vs *were* ratio (2293:70, or 32.7:1) provided strong evidence for interpreting the selection of *were* as motivated by hypercorrectness. Hypercorrect *were* was slightly more common in the IC (with 0.039 tokens pmw) than in the OC (0.032), and particularly so in US (and JM in the OC).

In order to identify open conditionals with hypercorrect *were*, a wide net had to be cast with the routine '. if P were …'. Subsequent manual processing of a subset of the output suggested that relevant tokens were vastly outnumbered by those with *was* and too infrequent to pursue intervarietal comparisons. They included (49) and (50):

(49) She held the straws in her hand, exposing the ends to the number requested. If she **were** asked for three, she held up three. If she **were** asked for four, she held up four. (US)

(50) Gordy was wary of releasing too many singles for fear of losing radio play. If he **were** worried about too many Tamla releases on the market at once, it still boggles the mind that this one would be withdrawn! (GB)

4.5 Hyperforeignism

'Hyperforeignism' arises from speakers misidentifying the distribution of morphological patterns found in loanwords, and wrongly extending them (Janda *et al.* 1994). The case for treating hyperforeignism as a category of hypercorrection is admittedly weaker than that for the other categories I have examined. While it is motivated by the misguided application of a 'rule', in this case involving loanword morphology, traditions of prescriptive censure are not in evidence (with the notable exception of *octopi*: see below). In this section I discuss three categories of hyperforeignism, all involving number in nouns, with many items yielding frequencies in GloWbE that were insufficient to pursue intervarietal comparisons.

4.5.1 Hypercorrect formation of -i final plural nouns from -us final singulars

Plural nouns with an *-i* suffix are sometimes formed from a *-us* final lexeme on the mistaken assumption that the latter is Latinate. The most well-known case is *octopi* as in (51), rather than *octopuses*, as the plural of *octopus* (which derives indirectly from Greek, rather than Latin, the Greek plural being *octopodes*). *Octopi* is censured by many prescriptive commentators, including Burchfield (1996: 316), who describes it as an 'oddity', and Allen, who regards it as a 'mistake' (www.grammarly.com/blog/octopi-octopuses/). In a paper based on the *HUGE Database of Usage Guides and Usage Problems* (see https://bridgingtheunbridgeable.com/hugedb/) Otto (2015) remarks that '[b]oth *octopi* and *octopodes* are usually proscribed, the latter because even though it is "good Greek", it sounds "pedantic"'.

(51) **Octopi** will become active, moving from one hide to the next. (AU)

A search of GloWbE revealed the OC to be more prone to the hypercorrect use of *octopi* (as indicated in its weak *octopuses* vs *octopi* ratio of 1.59:1) than the IC (with its stronger ratio of 3.29:1). Another example is the use of *platypi* as the plural of *platypus* as in (52), also criticised by some commentators. Like *octopus*, *platypus* is etymologically Greek despite its Latinate ending. In GloWbE there are 7 tokens of *platypi* (all in the IC varieties) and 62 of *platypuses*: 59 in the IC (including 38 in AU) and three in the OC.

(52) It can softly call other **platypi** with a soft, puppy-like growl/croon. (AU)

Further less common examples are *apparati*, *censi*, *foeti* and *prospecti*.

4.5.2 *Hypercorrect use of* -i *final singular nouns as plurals*

Singular *-i* final nouns are sometimes hypercorrectly used as plurals by analogy with Latinate plurals such as *fungi* and *cacti*. The highest-frequency cases are *Kiwi/kiwi* (fr. Maori 'a New Zealander', or 'a flightless bird that is native to New Zealand') as in (53), and *tsunami* (fr. Japanese 'a destructive sea wave') as in (54). Of the 60 hits for 'Kiwi/kiwi are' in GloWbE, 58 were from the IC varieties (and of these 50 are from NZ). Of the 48 tokens of plural *tsunami*, the majority were from the OC (25, or 0.38 pmw), rather than the IC (23, or 0.18 pmw).

(53) **kiwi** are feisty protectors of their nests and territory. (NZ)
(54) Wars, famines, floods, earthquakes, and **tsunami** are all disasters (MY)

Three further lower-frequency items are *taxi* (fr. French 'a car for public hire'), *Nazi* (fr. German 'a member of the National Socialist (German Workers') Party') and *yogi* (fr. Hindi 'one who practises yoga'). There are approximately three times more tokens of plural *taxi* in the OC (12, or 0.185 pmw) than in the IC (8, or 0.064pmw). There were 12 tokens of plural *Nazi* (all in the IC) and 4 tokens of plural *yogi*.

4.5.3 *Hypercorrect use of* -i *final plural nouns as singulars*

Plural *-i* final nouns are sometimes used as singulars in English, presumably by analogy with such *-i* final singular nouns as those discussed in section 4.5.2 above. Examples are *alumni*, *fungi* and *stimuli*.

(55) Comrade Olaitan a University of Lagos **alumni** is well-known (NG)
(56) Ringworm: This is caused by a **fungi**, not a worm. (AU)
(57) Tourism is recognized to be a **stimuli** for learning (GB)

In GloWbE there were 46 tokens (0.37 pmw) of singular *alumni* in the IC, as opposed to 62 (0.96) in the OC;14 tokens (0.17 pmw) of singular *fungi* in the IC, as opposed to 4 (0.07) in the OC; and three tokens (0.02 pmw) of singular *stimuli* in the IC, as opposed to two (0.03) in the OC.

5 Discussion and conclusion

The present corpus-based study complements more familiar sources of information on hypercorrection such as attitude elicitation studies, usage guide discussions and analyses in comprehensive grammars. The nature and extent of five categories of hypercorrection were investigated: case-marked pronouns, *-ly* adverbs, agreement with number-transparent nouns, (extended uses of) irrealis *were*, and 'hyperforeign' noun suffixation. The vast resources of GloWbE proved to be indispensable in determining whether particular usages qualify as instances of hypercorrection, the quantitative criteria employed being relative paucity of numbers vis-à-vis the more established standard counterparts, along with overall paucity of numbers in the corpus. In many cases the extent of prescriptive censure in usage handbooks provided further support

for a hypercorrectness interpretation (for example, for nominative pronouns in coordinative *between*-phrases, for -*ly* adverbs derived from flat adverbs such as *thusly*, and for 'pseudo subjunctive' *were* in non-conditional constructions).

Some notable varietal findings were recorded. One was a tendency for AmE to be more strongly associated with hypercorrection than BrE in a number of the categories, a finding no doubt attributable to the robustness of grassroots and institutional prescriptive traditions in the USA, which have seen AmE subjected to surveillance via grammar checkers, usage guides, freshman English textbooks and press columns written by 'language experts' (see further Drake 1977; Pullum 2009: Milroy & Milroy 2012). According to Leech *et al.* (2009: 264), 'Prescriptivism maintains it hold over written AmE through channels which are absent from the UK, such as handbooks for obligatory freshman English courses, and the pronouncements of "language mavens" in the press.' As an example consider the extensive attention given in American usage guides to the adverb *thusly*, whose greater prevalence and acceptance in AmE than in BrE are confirmed by GloWbE frequencies and by the findings of Lukač & Tieken-Boon van Ostade's (2019) attitude study. Another is the selection of a nominative pronoun as the complement of comparative *than* or *as*, an option far more popular in AmE than in BrE. The cross-varietal frequencies for this construction further indicated that some epicentral influence may emanate from AmE: as suggested by the strong showing in the OC of PhilE, a variety historically strongly influenced by AmE. Similar influence was also in evidence with other expressions, such as *feel badly*, frequencies for which were extremely high in AmE and CanE, but also relatively high in PhilE and JamE (Canada and Jamaica being countries which, like the Philippines, not only have historical connections with the USA, but also share geographical proximity with it). Curiously, despite evidence of a thriving complaint tradition in Australia and New Zealand – where the practice of publishing letters to the editor is more popular than it is in Ireland, the UK, the USA and Canada (Lukač 2018b: 8) – there was little evidence of co-patterning between the Antipodean varies, and only rarely did they display an appetite for hypercorrection stronger than that of the other IC varieties (AusE did so for nominative pronouns in coordinative constructions, and NZE for the use of relative *whom* as subject). A possible explanation is the relatively enlightened nature of the prescriptive tradition in Australia and New Zealand, informed by input from academic linguists (such as Pam Peters, author of the influential corpus-informed *Cambridge Australian English style guide* (1995) and *Cambridge guide to Australian English usage* (2007), and Kate Burridge, prominent Australian linguist and regular presenter of language segments on radio), along with the educational reforms referred to above (see further Severin 2017).

Several tendencies were found to be associated with the two varietal macro-groupings, the IC and the OC. A higher incidence of hypercorrection for the IC was noted in the case of relatively established constructions which have been the target of concerted prescriptive commentary, and particularly commentary transmitted over a long period of time via institutionalised channels. This was certainly so with nominative pronouns after *than* and *as* (the most frequent non-coordinative pronominal construction in the study),

prepositional phrases typified by *between you and I* (the most frequent coordinative construction), 'Type 3' *ly*-adverbs which overlap with their non-*ly* counterparts in specific collocations (the most frequent of the three -*ly* adverb categories), and to a lesser extent with pseudo-subjunctive *were* and hyperforeign plurals such as *octopi*. While in IC countries the prescriptive complaint tradition is focused on intralinguistic prestige-driven variation (with the exception of hyperforeignism), that which is found in OC countries tends not to be concerned with hypercorrectness, but rather aimed generally at local usages perceived to be L2 English errors that flout appropriate norms of correctness. While in some cases such 'errors' are associated with substrate languages (e.g. in SingE the use of passive constructions that can be traced to Malay (Bao & Wee 1999) and relative constructions from Chinese (Alsagoff 2001)), in other cases they involve features found across many OC varieties (e.g. the omission and insertion of articles and the use of the progressive with stative verbs).

Turning to the OC, it must be conceded that it was sometimes difficult to distinguish between hypercorrect usages on the one hand, and L2 errors or spurious neologisms on the other. Consider, for example, the dysfluent syntax in the coordinated verb object example in (16) above, and (58) below, where the aberrant syntax and the misuse of *allocation* might tempt one to analyse *longly* as an error rather than a hypercorrect variant of *long*.

(58) Stay in a bed too **longly** promotes on allocation and can be dangerous to your skin [TZ]

Having selected valid OC tokens, with due diligence, and having analysed their frequencies, I observed a tendency for hypercorrection in the OC to be associated with lower-frequency constructions. Given that most such constructions have not been singled out for traditional prescriptive censure, the question becomes: what factors might help us explain this tendency? In the case of some constructions, including *whom* used as the subject of an embedded clause, the complexity of the syntax might provoke more uncertainty amongst L2 English speakers in OC countries than amongst L1 English speakers in IC countries. Another factor that might be hypothesised to promote hypercorrection in OC varieties is what might be described as the insensitivity or unawareness of speakers to IC conventions and traditions. For example, *outrightly*, which was over eight times more frequent in the OC than in the IC, is recognisably unidiomatic for L1 speakers but may not be so for speakers in OC varieties.

In addition to dialect, register can sometimes be a factor that influences hypercorrection frequencies. Consider the tendency for *uprightly* to modify the verbs *walk* and *act* in theological and scriptural contexts where the -*ly* suffix arguably adds an archaic biblical dimension, a finding consistent with the strong occurrence of this adverb in countries with large Christian populations, such as the USA and Nigeria. Another example is the use of *muchly* in contexts of situation that involve lightheartedness or jocularity, as in:

(59) Thankies **muchly** for starting my Friday with big laughs:) (GB)

In conclusion, I express the hope this study will encourage others to pursue further work on hypercorrection, a fascinating but still under-researched topic. For example, little is known about the register distribution of hypercorrection, or directions of diachronic change in hypercorrection, topics whose empirical investigation will require the availability of suitable data from large corpora.

Author's address:

School of Humanities and Languages
University of New South Wales
Sydney NSW 2052
Australia
p.collins@unsw.edu.au

References

Aarts, Flor. 1994. Relative *who* and *whom*: Prescriptive rules and linguistic reality. *American Speech* 69, 71–9.

Allen, Robert (ed.). 1999. *Pocket Fowler's Modern English usage*. Oxford: Oxford University Press.

Alsagoff, Lubna. 2001. Tense and aspect in Singaporean English. In Vincent Ooi (ed.), *Evolving indentities: The English language in Singapore and Malaysia*, 79–88. Singapore: Times Academic Press.

Angermeyer, Philipp S. & John Victor Singler. 2003. The case for politeness: Pronoun variation in co-ordinate NPs in object position in English. *Language Variation and Change* 15, 171–209.

Bao, Zhiming & Lionel Wee. 1999. The passive in Singapore English. *World Englishes* 18, 1–11.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

Burchfield, Robert W. 1996. *The new Fowler's modern English usage*. Oxford: Oxford University Press.

Burridge, Kate. 2010. Linguistic cleanliness is next to godliness: Taboo and purism. *English Today* 26(2), 3–13.

Butterfield, Jeremy (ed.). 2015. *Fowler's dictionary of modern English usage*. Oxford: Oxford University Press.

Cameron, Deborah. 1995. *Verbal hygiene*. London & New York: Routledge.

Campbell, Lyle. 1998. *Historical linguistics: An introduction*. Cambridge, MA: MIT Press.

Collins, Peter. 2020. Comment markers in world Englishes. *World Englishes*. Early view. Published online 26 September 2020. https://doi.org/10.1111/weng.12523

Collins, Peter & Xinyue Yao. 2019. Colloquialisation and the evolution of Australian English: A cross-varietal and cross-generic study of Australian, British and American English from 1931 to 2006. *English World-Wide* 39(3), 253–77.

Corbett, Philip B. 2008. Subjunctivitis. http://afterdeadline.blogs.nytimes.com/2008/10/13/subjunctivitis/

Crystal, David. 2006. Into the twenty-first century. In Lynda Mugglestone (ed.), *The Oxford history of English*, 394–413. Oxford: Oxford University Press.

Curzan, Anne. 2014. *Fixing English: Prescriptivism and language history*. Cambridge: Cambridge University Press.

Davies, Mark. 2008–. *The Corpus of Contemporary American English*. Available online at www.english-corpora.org/coca/

Davies, Mark. 2010–. *The Corpus of Historical American English*. Available online at www.english-corpora.org/coha/

Davies, Mark. 2013. *The Corpus of Global Web-Based English*. Available online at www.english-corpora.org/glowbe/

Davies, Mark & Robert Fuchs. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36, 1–28.

Davies, Winifred V. & Evelyn Ziegler. 2015. *Language planning and microlinguistics: From policy to interaction and vice versa*. Basingstoke and New York: Palgrave Macmillan.

DeCamp, David. 1972. Hypercorrection and rule generalization. *Language in Society* 1(1), 87–90.

Drake, Glendon F. 1977. *The role of prescriptivism in American linguistics 1820–1970*. Amsterdam: John Benjamins.

Eckman, Fred R., Gregory K. Iverson & Jae Yung Song. 2013. The role of hypercorrection in the acquisition of L2 phonemic contrasts. *Second Language Research* 29(3), 257–83.

Fowler, Henry W. & Francis G. Fowler. 1906. *The King's English*. Oxford: Clarendon Press.

Garner, Bryan. 1998. *A dictionary of Modern American usage*. New York: Oxford University Press.

Gowers, Sir Ernest. 1965. *Fowler's modern English usage*. Oxford: Oxford University Press.

Huddleston, Rodney & Geoffrey Pullum *et al*. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hundt, Marianne. 2009. Normverletzungen und neue Normen. In Marek Knopka & Bruno Strecker (eds.), *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*, 117–40. Berlin: Walter de Gruyter.

Janda, Richard, Brian Joseph & Neil Jacobs. 1994. Systematic hyperforeignisms as maximally external evidence for linguistic rules. In Susan Lima, Roberta Corrigan & Gregory Iverson (eds.), *The reality of linguistic rules*, 67–92. Amsterdam: John Benjamins.

Kachru, Braj. 1985. Standards, codification and sociolinguistic realism: The English language in the Outer Circle. In Randolph Quirk & Henry Widdowson (eds.), *English in the world: Teaching and learning the language and literatures*, 11–36. Cambridge: Cambridge University Press.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: The Center for Applied Linguistics.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.

Leonard, Sterling A. 1929. *The doctrine of correctness in English usage, 1700–1800*. Madison, WI: University of Wisconsin Press.

Loureiro-Porto, Lucia. 2017. ICE vs GloWbE: Big data and corpus compilation. *World Englishes* 36, 448–70.

Lowth, Robert. 1762. *A short introduction to English grammar*. London: A. Millar, R. and J. Dodsley.

Lukač, Morana. 2018a. *Grassroots prescriptivism*. University of Leiden dissertation. https://scholarlypublications.universiteitleiden.nl/handle/1887/67115

Lukač, Morana. 2018b. Grassroots prescriptivism: An analysis of individual speakers' efforts at maintaining the standard language ideology. *English Today* 34(4), 5–12.

Lukač, Morana. 2018c. What is the difference between *thus* and *thusly*? *E-Rea: Revue électronique d'études sur le monde anglophone*. https://doi.org/10.4000/erea.6152

Lukač, Morana & Ingrid Tieken-Boon van Ostade. 2019. Attitudes to flat adverbs and English usage advice. In Sandra Jansen & Lucia Siebers (eds.), *Processes of change: Studies in Late Modern and Present-Day English*, 159–81. Amsterdam: John Benjamins.

Mair, Christian. 2006. *Twentieth-century English: History, variation, standardization*. Cambridge: Cambridge University Press.

Merriam-Webster's dictionary of English usage. 1994 (repr.). Springfield, MA: Merriam-Webster.

Milroy, James & Lesley Milroy. 2012. *Authority in language: Investigating language prescription and standardization*, 4th edn. London: Routledge.

Mittins, W. H., Mary Salu, Mary Edminson & Sheila Coyne. 1970. *Attitudes to English usage: An enquiry by the University of Newcastle upon Tyne Institute of Education English research group*. Oxford: Oxford University Press.

Morris, William & Mary Morris. 1975. *Harper dictionary of contemporary usage*. New York: Harper & Row.

Nevalainen, Terttu. 1997. The processes of adverb derivation in Late Middle and Early Modern English. In Matti Rissanen, Merja Kytö & Kirsi Heikkonen (eds.), *Grammaticalization at work: Studies in long-term developments in English*, 145–89. Berlin: Mouton de Gruyter.

Odlin, Terence. 1989. *Language transfer: Crosslinguistic influence in language learning*. Cambridge: Cambridge University Press.

Otto, Inge. 2015. A fuss about the octopus. *English Today* 31(1): 3–4. https://doi.org/10.1017/S0266078414000479

Oxford English Dictionary Online. 2020. www-oed-com

Partridge, Eric. 1963. *Usage and abusage: A guide to good English*. Baltimore: Penguin Books.

Peters, Pam. 1995. *The Cambridge Australian English style guide*. Melbourne and Cambridge: Cambridge University Press.

Peters, Pam. 2007. *The Cambridge guide to Australian English usage*, 2nd edn. Melbourne and Cambridge: Cambridge University Press.

Pickett, Joseph P., Steven Kleinedler & Susan Spitz (eds.). 2005. *The American Heritage guide to contemporary usage and style*. Boston, MA: Houghton Mifflin.

Pullum, Geoffrey. 2009. 50 years of stupid grammar advice. *The Chronicle of Higher Education*, 17 April.

Quinn, Heidi. 2005. *The distribution of pronoun case forms in English*. Amsterdam: John Benjamins.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Schneider, Edgar. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Severin, Alyssa A. 2017. Vigilance or tolerance? Younger speakers' attitudes to Australian English usage. *Australian Journal of Linguistics* 37(2), 156–81.

Straaijer, Robin. 2014. *Hyper usage guide of English*. http://huge.ullet.net/

Tieken-Boon van Ostade, Ingrid. 2020. *Describing prescriptivism: Usage guides and usage problems in British and American English*. London: Routledge.

Touchie, Hanna Y. 1986. Second language learning errors: Their types, causes, and treatment. *JALT Journal* 8(1), 75–80.

Wolfram, Walt. 1991. *Dialects and American English*. Englewood Cliffs, NJ: Prentice Hall.