

On being *happier* but not *more happy*: Comparative alternation in speech data

Alexandra D’Arcy
University of Victoria
adarcy@uvic.ca

English adjective comparison is increasingly the focus of corpus linguistic research, but it is much less studied in the variationist framework. These two traditions converge, however, in revealing robust variation between historical inflection (*happier/happiest*) and newer periphrasis (*more/most happy*). However, our understanding of the strategies for comparison comes from written genres. In contrast, very little is known about comparison in vernacular speech. Since periphrastic comparison emerged as a change from above, the lack of spoken evidence proves a critical gap in our knowledge. To address this gap, this paper examines comparison strategies in New Zealand English, drawing on the whole of the Origins of New Zealand English Archive (Gordon et al. 2007). Analysis of 1400 tokens reveals a striking result. Consistent with reports elsewhere, inflection is the preferred mode of comparison. However, consideration by lexical item reveals a system that is not, in fact, variable. Rather, across the history of this variety (speakers born 1851-1982), individual adjectives pattern one way (inflection) or the other (periphrasis); in speech, the form of comparison has consistently been lexically conditioned, and by extension, invariant. This paper explores a number of explanations (e.g. variation is genre-specific or variety-specific, or may only be visible in extremely large corpora), and ultimately concludes that in speech, historical variation resulted in the full ‘regularization of a confused situation’ (Bauer 1994:60).

Keywords: adjective comparison; vernacular speech; variation; New Zealand English

1 A showcase of grammatical variation

By all accounts, comparative alternation in English is robust and long established as a variable feature of the language. Historical inflection, as in (1a,c) competes with ‘innovative’ periphrasis (1b,c), a layering of forms that has been attested since the Middle English period (Pound 1901; Mitchell 1985; Kytö 1996). This competition may also result in the combination of the two strategies, creating a hybrid such as is exemplified in (1d). This final strategy, an unmitigated minority option in Present Day English, was in Middle and Early Modern English a viable option for comparison, both comparative and superlative (Kytö 1996:124; see also Schlüter 2001; Kytö & Romaine 2006).

- (1) a. *Subtler* and more successful methods are sometimes tried.
(WWC/G45/124)¹
- b. Sondra ...was *more subtle*. (WWC/K87/121)
- c. It was the *most silly*-- it was the *silliest* thing I've ever heard of.
(V. Sheehy, b.1896)
- d. You know the *most silliest* things you know. (fyn01-5b)

Although linguistically (i.e. referentially) equivalent, however, inflection and periphrasis are not in all instances socially (i.e. ideologically) equivalent. This lack of parity is most clear in the context of second language learning, but it also finds voice in less formalized metadiscourse, both online and otherwise (e.g. Urban Dictionary). While the former likely arises from pedagogical imperatives to teach 'standard' language practice, the latter is rooted in the nuances of communicative competence, determined by community-internal norms and sociolinguistic cultures. Regardless, it is clear that adjective comparison is an area of the grammar that is rife with notions of 'right' and 'wrong'.

Consider what seems a straightforward example, an adjective such as *happy*. On a language forum (WordReference.com; November 27 2007), use of the periphrastic comparative in 'They are more happy than the mean people' spurred two threads, both of which derided the construction in favour of inflection. Ultimately, *more happy* was deemed colloquial and better avoided in a range of formal situations. An English second language speaker concludes one of the threads with the assertion that 'for someone whose english is not the mother tongue, [...] it's worth not using it.' On a different forum (Yahoo!Answers; January 2011), a user asked which of the two options in (2) 'make [sic] more sense'. The answer, prescriptive in tone, refutes (2b): both 'make sense', but *happier* is 'correct' and *more happy* 'is not'.

- (2) a. I'm happier than ever before.
- b. I'm more happy than ever before.

Metadiscourse of this kind is particularly insightful from a sociolinguistic perspective. As with other rule-based, prescriptive ideologies of language usage, commentary typically is not reflected in practice (see also D'Arcy & Tagliamonte 2010). Thus, despite overt notions of correct and incorrect usage (e.g. *when a word ends with <y>, drop the <y> and add <ier>*), a simple Google search quickly reveals not only a plethora of variation, as illustrated by the examples in

¹ The parenthetical information following examples from corpora provides details of provenance. For examples from the Wellington Corpus of Written New Zealand English (WWC), the category and transcript number and the individual line number are given; for examples from the Origins of New Zealand English Archive (ONZE), the speaker details are provided. This takes the format of either a name or a code, depending on the protocol of the collection from which the example is drawn.

(3), but also creative and seemingly willful manipulation of the choice between inflection and periphrasis.

- (3) a. The first step to becoming *more happy* is to realize you don't know how to make yourself *happier*. (Happy Deviant; Positive psychology: How to construct a happier life – Part 1; May 15 2011)
- b. When asked if they'd like to be *more happy* the majority of people shout 'yes' because 'who wouldn't want to be *happier*?' (Jodi Lee – Life Designer; Can you recognize what makes you happy?; November 5 2010)
- c. *Happier* people tend to be healthier. The *more happy* you feel, the less prone to illness you will be. (Anti-Aging Web Magazine; How growing old makes you happier!)
- d. I was *happy* with my stock '08 SG [...] added a 6" long ride shield and was *more happy* [...] added aileron grips and again I was *more happy* than ever [...]. If I get any *happier* I don't think I'll be able to live with myself. (HDForums.com; Happy to be more happy (happier?); November 15 2009)

The examples in (3), all of which evince variation within connected stretches of writing, provide illustration of the robust nature of the choice mechanism that is operative in adjective comparison. They also illustrate the prosaic possibilities that this variation enables. In (3c), for example, the deployment of both the inflectional and the periphrastic constructions derives in each case from the stylistic drive for structural symmetry (*happier...healthier; the more...the less*). In other words, the language itself allows overt manipulation of comparative forms when it would be entirely possible (indeed, 'correct') to use only inflectional comparison for adjectives such as *happy*. To this end, the example in (3d) is striking not only for the variation in the text but also for the title of the post, which openly plays with the alternation between comparative strategies. As with (3c), the author opts for symmetry (in this case, bracketing the title with identical elements, *happy...happy*) while at the same time highlighting the fact that another option exists.

We therefore have a situation whereby adjective comparison is the subject of metalinguistic discussion among language learners, grammarians and language mavens, and the different strategies can be overtly manipulated to achieve a range of stylistic and pragmatic effects. Linguists tend to target known features, those that are treated in language grammars and by other researchers (Cheshire 1999). It therefore follows that a large body of academic work targets comparison. From a paradigmatic perspective, much of this is based in the corpus linguistic tradition (e.g. Bauer 1994; Kytö 1996; Kytö & Romaine 1997, 2000, 2006; Leech & Culpepper 1997; Mondorf 2003; etc.), the results of which suggest that comparative alternation is a robust and well-attested phenomenon. If

scholarly output is any evidence, however, then it is clear that variationists have been much less keen to investigate this feature (cf. Hilpert 2008; Scrivner 2010).

That the variationist tradition has not weighed in on comparative alternation is striking. By all accounts, variation between inflection and periphrasis is highly constrained. All levels of linguistic analysis are implicated (phonology, morphology, syntax, semantics, pragmatics, the lexicon, etc.), a fact which led Mondorf (2009:1) to describe this feature as ‘a showcase of grammatical variation’. In sum, comparative alternation presents a complex phenomenon, and so it seems (intuitively anyway) to be an ideal candidate for variationist analysis, particularly if one is interested in mechanisms and pathways of change.

2 Historical perspective

Competition between inflection and periphrasis has been operative in English comparison since roughly the thirteenth century (Pound 1901; Mitchell 1985; Kytö 1996; Kytö & Romaine 1997, 2000, 2006). In other words, the spread of periphrasis within the sector constitutes a longstanding change with modern-day reflexes. However, the diachronic picture is atypical because adjective comparison does not exhibit the normal trajectory of replacement (i.e. $A > B$).

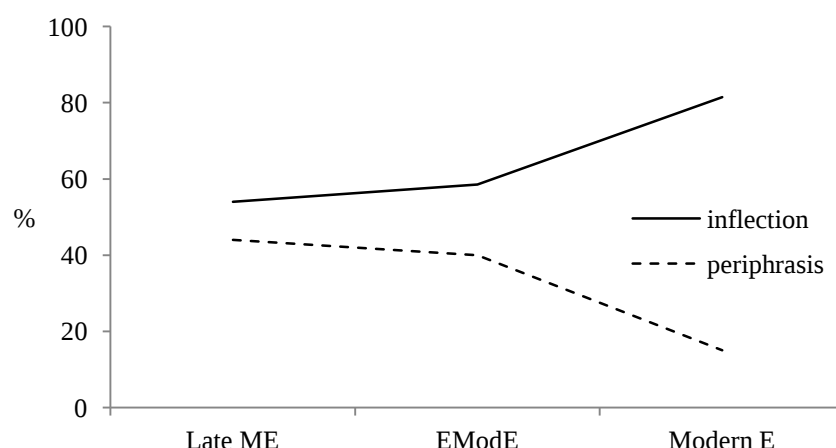


Figure 1. Historical trajectories of inflection and periphrasis, comparative (from Kytö & Romaine 1997:336, Fig.1)

Historically, the periphrastic forms are innovations; Old English of course was inflectional. When the periphrastic option for comparison first developed, it diffused relatively slowly until approximately the end of the fourteenth century, when it began to increase quite steadily (Mitchell 1985). This initial trajectory follows the established template for innovations, which generally entail introduction, incipient variation, and subsequent increase (i.e. the S-curve of

linguistic change; Weinreich et al. 1968; Bailey 1973; Altmann et al. 1983; Kroch 1989; Labov 1994; though see Denison 2003). However, periphrasis did not continue to spread throughout the sector, systematically pushing out or marginalizing inflection. Instead, periphrasis seems to have peaked during the Late Middle English period (Pound 1901); since that point, as illustrated in Figure 1, inflection has been busily ‘reasserting itself’ (Kytö & Romaine 2000:172). Contemporaneously, the majority of comparison (both comparative and superlative) is inflectional, with present day distributions having been achieved during the Late Modern English period (i.e. post 1710; Kytö & Romaine 1997).

The vast majority of our understanding concerning the history, development, and patterning of comparative alternation in English is based on evidence from written genres. From a historical perspective, this is perfectly valid; it is also a pragmatic necessity. However, it is also important to consider the mechanism by which periphrasis entered the field of comparison. The answer is particularly germane to the question of variation and change, since changes from above (adaptive) have sociolinguistic properties that are distinct from those associated with changes from below (evolutionary).

Despite the evolutionary tendency for English to be moving from synthetic to periphrastic syntax, it has been suggested that periphrastic comparison is not a language-internal development. Instead, it likely emerged under the prestige influence of Latin primarily, but also, to a lesser extent, French (Mustanoja 1960:279). In other words, periphrasis in this instance is a historical change from above. In this case then, what happens in speech—and in unscripted, casual, vernacular speech in particular—presents a potentially rich source of empirical evidence to add to our understanding of this grammatical feature.

3 Data and method

The data for this study come from the Origins of New Zealand English Archive (ONZE), one of the largest repositories of longitudinal spoken English data available for sociolinguistic analysis. Consisting of three collections (the Mobile Unit, the Intermediate Archive, and the Canterbury Corpus), ONZE includes over 1000 hours of casual speech from more than 700 individuals, covering the history of New Zealand English, 1850 to the present (for full details, see Gordon et al. 2004 and Gordon et al. 2007).

Following the *principle of accountability* (Labov 1966, 1972), the cornerstone of variationist methodology, the recordings were exhaustively searched and all instances of comparison (comparative and superlative) were extracted. All told, these materials include 2621 tokens, 1221 suppletive (e.g. *good, better, best*) and 1400 non-suppletive (e.g. *nice, nicer ~ more nice, nicest ~ most nice*). The focus of this analysis is the non-suppletive group, as this is the site of variation; suppletive comparison is invariant.

The details of the analysis are provided in Table 1, which provides the breakdown by collection within the Archive.

Table 1. ONZE sample and data (non-suppletive adjectives only)

<i>corpus</i>	<i>speaker years of birth</i>	<i>N speakers</i>	<i>N tokens</i>
Mobile Unit	1860-1919	32	197
Intermediate	1891-1963	56	449
Canterbury Corpus	1922-1982	151	754
total N		239	1400

4 Results

In terms of overall results, the data conform to expectation in that the majority of comparison is inflectional (68.5%). They are also typical in that inflection is more frequent in the comparative (71.9%, $N = 983$) than it is in the superlative (60.4%, $N = 417$). Moreover, even though the data represent colloquial speech, hybrid forms such as (1d) are exceptionally rare: They account for less than 1% of the data overall (0.79%; $N = 11$).

The picture presented by these aggregate results is thus consistent with that presented by analyses based on written data. During the Modern Period, comparison is largely (but far from exclusively) an inflectional phenomenon, and conflation of the strategies is rare. However, given that the ONZE materials span more than a century in apparent time, and because they capture the history of the variety, they have the potential to be insightful with respect to the establishment of the present-day New Zealand system.

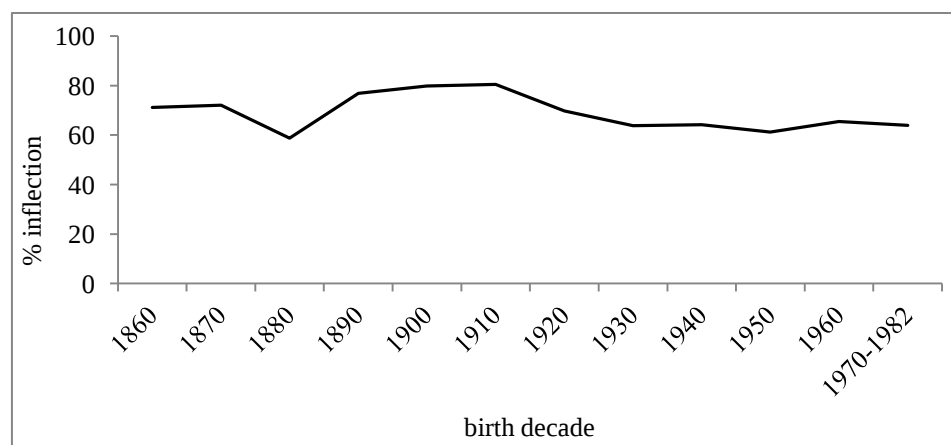


Figure 2. Overall distribution of inflectional comparison across time in ONZE

If the temporal dimension of the data is examined, as in Figure 2, the distributional workload appears stable across time. Indeed, there is no correlation between time (as a factor of speaker birth year) and the overall frequency of inflection in these data ($r = -0.325632$, $p = .167739$). This result strongly suggests that comparative alternation was fixed prior to the formative period of New Zealand English; the variety underwent no distributional re-organization

subsequent to permanent British settlement in the colony. Thus, despite the diachronic depth of these speech data, time is not implicated in the variation whatsoever. As it happens, there is another critical respect in which this is the case as well. This point will be illustrated shortly. First, however, there are grammar-internal factors to explore, factors which may constrain the degree of comparative variation that is possible within any given dataset.

A key aspect of adjective comparison is that the window for variation is fairly circumscribed: Not all adjectives may be compared using both inflection and periphrasis. This is because over time, the modes have specialized in certain linguistic environments. Ultimately, periphrasis has successfully ousted the older inflectional means from some contexts, but in others it is the inflectional type that has triumphed. For example, adjectives consisting of four or more syllables categorically require periphrasis (e.g. *academic*, *democratic*, *profitable*, *apologetic*, *enthusiastic*, *sophisticated*). The same is largely true of trisyllabic adjectives as well. Thus, a form such as *important* can only be modified as *more* or *most important*; **importanter* is not grammatical (descriptively) in native, unmonitored, adult language. This is quite distinct from the behaviour of adjectives such as *happy*, where both modes of comparison are possible and ‘grammaticality’ is a matter of prescription (i.e. the structure of the language allows variation).

This kind of categorical patterning has ramifications for variationist analysis, since the aim is to focus on those forms for which variation is possible. In the ONZE materials, there are more than 130 types of multi-syllabic adjectives that categorically take periphrasis to mark comparison. Such tokens are thus of little interest to a discussion of variation.

Where the bulk of variation putatively occurs is among bisyllable adjectives (e.g. *bolshy*, *clever*, *deadly*, *mature*, *narrow*, *pleasant*, *quiet*, *recent*, *wealthy*, *vivid*, *yellow*, etc.). Certainly with forms such as *bolshy*, *clever*, and *deadly*, inflection or periphrasis are both acceptable. However, there are forms that cannot vary. With adjectives such as *alive*, *carefree*, *complex*, *correct*, *human*, *nervous*, *open*, *passive*, *peaceful*, *private*, *senior*, *unfair*, *unjust*, *upset*, and *useful*, for example, inflection is not possible. As with multi-syllabic adjectives, the sole grammatical mode of comparison is periphrasis. What is particularly notable about the exceptional forms in the bisyllabic category, however, is that structurally (syllabically and segmentally), they are not distinct from those that are variable: As outlined in Figure 3, a large proportion of coda or final segments (a putative condition for inflection or periphrasis, cf. /-i/) overlaps the categories. This is important because it means that the determinants of (non) variability are not predictable. Having invariant adjectives also means that the window of variation is further restricted: The focus for quantitative modelling are the forms that can alternate, not those that cannot and which, by extension, do not.

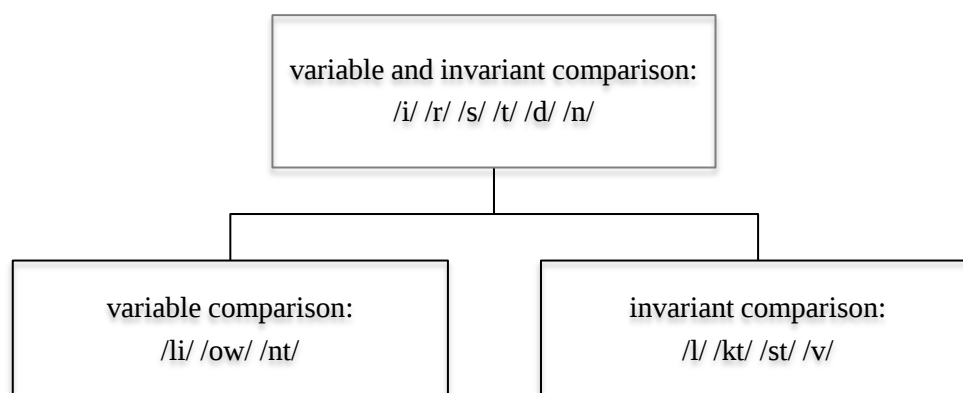


Figure 3. Coda or final segments of bi-syllabic adjectives in ONZE, according to mode of comparison

The consequence of categoricity is that tokens for which comparison is achieved strictly via inflection or strictly via periphrasis must be excluded from the analysis. In the ONZE materials, this category contains 838 tokens. Put in broader perspective, invariant comparison accounts for 60% of non-suppletive adjectives in the dataset. That is a huge proportion of forms that are simply not able to evince variation. Moreover, of the initial 1400 tokens, only 562 tokens remain in the analysis. These forms can alternate—at least, in principle they can.

As it happens, in these unscripted, casual speech data, there are contexts which, despite being *able* to vary, do not. Most notable among these are trisyllabic adjectives. In the normal case these are compared using *more* or *most*, but there are certain adjectives (e.g. *beautiful*, *dangerous*, *expensive*) for which cases of inflection are imaginable and acceptable. Indeed, such instances are easily located on the Web (e.g. ‘The online world is becoming *dangerouser* and *dangerouser*’, http://s1.zetaboards.com/H_P_Hummingbird/topic/3809578/1/). In speech-based ONZE, however, comparison of trisyllabic adjectives is strictly periphrastic.

At the other end of the syllable scale are monosyllabic adjectives; the ‘rule’ for these is that comparison is inflectional. Of course, there is a huge number of single-syllable adjectives that take variable comparison. Consider just the *bs*: *blue*, *blunt*, *bold*, *brash*, *brave*, *bright*, *broad*, etc. Despite being able to host both inflectional and periphrastic comparison, in ONZE these forms are compared ‘as they should be’, that is, inflectionally. Indeed, at over 95% ($N = 427/449$), the rate is nearly categorical. There is simply no variation in this group either.

Monosyllabic adjectives are also the largest syllable type within the dataset; they account for almost 80% of the remaining data. There is no reason to suspect that this distribution is atypical and that other analyses have not largely been based on monosyllabic forms as well. If such is the case, then it is these forms that are driving the higher overall rates of inflection that are regularly reported in the literature (e.g. Kytö & Romaine 1997, 2000). Note that this generalization holds within the full ONZE dataset as well, where monosyllabic

adjective comprise 68.7% of tokens (N = 962/1400) and 98.2% of inflection (N = 942/959). Moreover, the New Zealand results suggest that the nearly categorical status of inflectional comparison on monosyllabic adjectives is well entrenched: There is zero variability across time in ONZE. *Cheap* is always *cheaper*, never *more cheap*. *Dark* is always *darker*, *fast* is always *faster*, *hard* is always *harder*. This is a dataset that extends to the mid 19th century. For a form such as *cheap*, for example, there are tokens from speakers born in the period from 1893 through to 1980 but there is not one instance of periphrastic comparison (N = 45); *cheaper* or *cheapest* are the only forms attested.

In fact, once individual forms are considered, it quickly becomes apparent that there is no variation within lexical items in ONZE. The complete list of forms that exhibit variable comparison is given in (4): *clever* (N = 3), *common* (N = 2), *cool* (N = 8) and *silly* (N = 3). Of these, only two vary within a paradigm: *clever* in the comparative (4a) and *silly* in the superlative (4d). With the exception of (4d) (a correction), none varies either within a single utterance or within the speech of any single individual. In other words, there is little evidence that this variation is regular or systematic (i.e. it may be idiosyncratic).

- (4) a. They were *cleverer* than we were. (W. Oliver, b.1907)
 I was *more clever* at that than anything else. (A. Shacklock b.1891)
 b. The *commonest* type of nickname they got was... (J. Marin, b.1900)
 It seems to be getting more and *more common*. (L. Algie, b.1926)
 c. Sam was a little *more cool*. (mop94-4)
 South Intermediate, which was the *coolest* school ever... (fyn00-7)
 d. It was the *most silly*-- it was the *silliest* thing I've ever heard of.
 (V. Sheehy, b.1896)

Given that lexical items do not vary in this dataset, the complex, interwoven net of constraints that has been reported in the literature becomes problematic. Syllable structure, phonological structure of the root-final coda, stress, haplology, complement type, semantic status, and end-weight are just some of the constraints that have been reported to operate on comparative alternation (Kytö 1997; Leech & Culpepper 1997; Lindquist 1998, 2000; Mondorf 2003, 2009; Hilpert 2008, etc.). However, where these studies report robust variation, the ONZE materials do not provide any substantive evidence for it. These factors are thus neither relevant nor applicable in ONZE.

5 Resolving an analytical conundrum

These materials thus present an analytical conundrum. Contrary to other analyses of English comparative alternation, individual adjectives are invariant in their comparative strategy. This is consistent across the ONZE Archive, in both temporal and social space. There are three possible explanations: variety, corpus size, and genre or register.

5.1 Variety

The current data are from New Zealand English. Other research is based on either British English or American English. It is possible that New Zealand English is distinct from other World Englishes—and from the two primary Inner Circle varieties in particular—in not exhibiting (robust) comparative alternation. As the youngest of the colonial varieties, for example, New Zealand English has most recently undergone dialect formation. That such is the case allows for the possibility that when it coalesced as a distinct regional variety, the variation that characterizes adjective comparison in British English levelled (on *levelling* in new dialect formation, see Trudgill 2004).

One way to test for the role of language variety is to examine another corpus of New Zealand English data. One such source is the Wellington Corpus of Written New Zealand English (WWC; Bauer 1993). This is a 1-million-word corpus, modelled on the Brown and Lancaster-Oslo-Bergen (LOB) corpora; it contains 500 2,000-word samples from a range of written genres.

The most comprehensive variationist analysis of English comparative alternation is that of Hilpert (2008), who drew on the British National Corpus (BNC). Hilpert provided a list of 247 adjective types that alternated in the BNC materials. Using that list, the WWC was systematically searched, resulting in 1149 tokens of adjective comparison (comparative and superlative), compiled from 113 adjective types. The data are qualitatively different from those extracted from ONZE: They are variable at the level of individual lexical items—in some cases, quite robustly.

Table 2 presents a representative sampling of the WWC data in terms of raw occurrences (right), including a comparison with Hilpert's (2008) BNC findings (left). Most notable about the comparison is the similarity between the two sets of results with respect to the tendencies exhibited by individual forms.

The first five adjectives in the table occur in relatively high numbers in the WWC, and they are invariant. As such, it is tempting to conclude that New Zealand English is exceptional with respect to adjective comparison (i.e. that variation is not a feature of the variety). But such a conclusion would be premature. The token counts for these particular adjectives are extremely high in the BNC. The least frequent has a raw occurrence rate of over 1500. Of these adjectives, however, the most 'robustly' variable is *broad*, which takes periphrasis at a rate of just 0.4%. For all intents and purposes, these adjectives—despite being categorized as variable—pattern categorically. In other words, for forms such as these, variation is highly infrequent and exceptional.

In contrast, variation is very much in evidence for the forms illustrated in the bottom half of Table 2, both in the BNC and, most notably, in the WWC. Moreover, the New Zealand data exhibit variation despite the low frequencies at which these individual forms are attested in the WWC. This result contrasts starkly with that from ONZE, where these very same adjectives are categorically inflected (i.e. invariant).

Table 2. A sample comparison of results from the BNC and the WWC

Adjective	BNC		WWC	
	-er	more	-er	more
<i>big</i>	4466	1	39	0
<i>broad</i>	1588	7	12	0
<i>great</i>	15936	1	188	0
<i>hard</i>	1745	1	30	0
<i>small</i>	8816	5	62	0
<i>just</i>	9	8	2	1
<i>mature</i>	14	141	1	2
<i>rare</i>	231	22	3	1
<i>shallow</i>	125	6	6	2
<i>simple</i>	1115	60	8	1
<i>stable</i>	9	86	1	1
<i>subtle</i>	114	339	4	4
<i>sweet</i>	157	2	2	1

To summarize, the results once again present a confound. On the one hand, ONZE provides data from over 120 years of spoken New Zealand English and no variation at the level of the lexicon is evident: Individual adjectives pattern one way (inflection) or the other (periphrasis). On the other hand, the WWC provides data from just 4 years of written New Zealand English and among individual adjectives there is fairly robust variation (e.g., *rare*, *shallow*: 75% inflection; *sweet*, 66% inflection; *subtle*, 50% inflection).

This finding *ipso facto* rules out variety. New Zealand English does not have special status with respect to adjective comparison. Written New Zealand data exhibit the same variation as do written British data and written American data.

5.2 Corpus size

The next hypothesis to consider in explaining the anomalous New Zealand findings concerns corpus size. ONZE, while one of the largest extant English speech corpora, is nonetheless a significantly smaller corpus than what is generally used when investigating adjective comparison (the BNC, for example, contains 100 million words). Although 1400 non-suppletive tokens were initially extracted from ONZE, perhaps thousands more are required to uncover variation. This explanation, however, is not supported by the evidence. At 1 million words, the WWC is but one-hundredth the size of the BNC. Despite being radically smaller than the BNC, the WWC provides clear evidence of variability. It therefore cannot be that ONZE is too small to capture variation. There are over 1000 hours of unscripted speech in the Archive, and, at over 1.4 million words, ONZE is in fact a considerably larger corpus than is the WWC.

5.3 Genre

The remaining difference between the ONZE data and the data used in the vast majority of research on English comparative alternation concerns genre. ONZE is a spoken corpus, which is part of what makes it such a rich resource in English historical sociolinguistics. With one notable exception, to which I return shortly, all other work on comparative alternation has considered written evidence. Some details are provided in Table 3.

Some of these works provide a list of the adjectives used in the analyses; many do not. Hilpert's (2008) contribution is particularly valuable because it includes not only a listing of the token types but also a break-down of these types by comparative mode, as either inflectional or periphrastic. As discussed above, many of the adjectives that were included in his analysis as variable were, at best, only marginally so (standard variationist methodology sets the cut-off point for categorical behaviour at $\leq 95.0\%$; see Guy 1988). As it happens, however, this is particularly true of the data used in Scrivner (2010). This is a critical point, because Scrivner (2010) is the only other study to have focused exclusively on speech. Her data were drawn from the spoken component of the Corpus of Contemporary American English (COCA), specifically, from those sections that contained unscripted speech. These consisted of talk shows and news programs (e.g., *Good Morning America*, *The Today Show*, *60 Minutes*, etc.).

Table 3. English datasets used in previous analyses of comparative alternation

Work	Corpus
Bauer 1994	The Times, The New York Times
Kytö 1996	Helsinki Corpus
Kytö & Romaine 1997	ARCHER, BNC
Kytö & Romaine 2000	Corpus of Early American Texts, ARCHER
Kytö & Romaine 2006	CONCE
Mondorf 2009	newspapers, fiction, BNC
Hilpert 2008	BNC
Scrivner 2010	COCA

Identical to the methodology used by Hilpert (2008) and following standard variationist practice, Scrivner (2010) included only adjectives that alternated modes of comparison. In the spoken subset of COCA mined by Scrivner, there were 90 such types. However, of these 90 variable adjectives, a full 43% (representing 39 types) exhibited variability at levels below 5%. In fact, most were variable well below this level. Crucially, this is not an issue of token numbers. The least frequent of these adjectives occurs 33 times, but most of the virtually categorical forms occur at frequencies in the hundreds and even thousands.

Within the variationist paradigm, distributions above the 95%/5% threshold are considered 'nearly categorical', 'exceptional', and 'statistically problematic' (see, e.g., Guy 1988). For this reason it is standard practice to remove them from

the quantitative model. The concern here, however, is not statistical validity but something much more fundamental. A non-trivial proportion of Scrivner's data is 'virtually categorical'; variation is hard to find. What I would like to suggest, therefore, is that variation in adjective comparison is crucially affected by genre. In writing it remains a robust phenomenon, but in speech, it is marginalized. Such a hypothesis accounts not only for the marked differences between the ONZE results and those of previous analysis of comparative alternation in English, it also accounts for the stark contrast between the findings from ONZE and those from the WWC.

6 Discussion and conclusion

Speaking of comparative alternation, Bauer (1994:60) once suggested that 'change' in the 20th century was not a matter of strategy *per se* but of the 'regularization of a confused situation'. Ultimately, comparison became 'more predictable': Periphrasis and inflection specialized. The added insight provided by the ONZE data (and, arguably, by the COCA data as well) is that nowhere is this more apparent than in speech.

In the normal, unmarked case—in speech as in writing—adjectives with four or more syllables require periphrasis. In speech, however, this is extended to adjectives with three syllables. The category of unmarked comparison is also extended in speech to include inflection on monosyllabic forms.

What of bisyllabic adjectives? These are the bastions of variation in writing, where stylistic factors such as symmetry and word play, pragmatic factors such as the desire to be witty, and structural factors such as syntactic, semantic, and pragmatic complexity all exert their effects. This is uncontentious. In speech, however, comparison has regularized. What is notable about bisyllabic adjectives is that while generalizations can be made about phonological effects (for example, final /l/ favours periphrastic comparison), these generalizations do not apply across the board. Exceptions exist (e.g. *narrower* vs. *more mellow*). This is strongly suggestive that the effects are lexical (likely deriving from frequency effects; see, e.g., Braun 1982; Quirk et al. 1985; Hilpert 2008). They are not predictable on structural grounds.

Indeed, this raises the question, unasked before now, of the extent to which speech was ever variable. The answer is unclear, but ONZE, a rare source of diachronic evidence, provides potentially rich insight. As summarized by McCarthy (1991:143f):

We do not know enough about the acceptable norms of grammar in speech since, up to now, our grammar books have been largely formulated from introspective and written data. A good grammar of spoken English might well contain a few surprises.

The written language may have always been more variable, particularly if periphrastic comparison entered as a change from above. To that end, consider

another well-known case of a prestige borrowing, the *wh*- relative pronouns. Romaine (1982:212) made the now famous observation that ‘the infiltration of *wh*- ... can be seen as completed in the modern written language ... but it has not really affected the spoken language.’ A similar argument could be made for comparative alternation. In this case, it is not that the periphrastic forms have not infiltrated speech. Clearly they have. Rather, for speech the options have specialized, conditioned by individual adjectives. The fall-out from this change from above remains visible in writing, where variation abounds for mono- and bisyllabic adjectives in particular, but it is much less of a factor for these same forms in speech.

To conclude, the possibility for variation in adjective comparison is not ruled out in speech. Instead, I would like to suggest that it be recognized for what it is: The exception rather than the norm, a locus of variation that is not, in fact, particularly variable. In writing it is possible to be *more happy*, but in speech, it is generally the case that *happier* is the way to be.

Acknowledgements

I gratefully acknowledge the members of the Origins and Development of New Zealand English team at the University of Canterbury, who prepared, transcribed, and time-aligned the ONZE materials. I thank Alesia Malec, my Research Assistant on this project, whose position was funded by the University of Victoria Work Study Program (Student Awards and Financial Aid). Appreciation is also extended to Julie Auger, Heidi Quinn, and Olga Scrivner for their help at various stages in the conduct of this research.

References

- Altmann, G., von Buttlar, H., Rott, W. & Strauß, U. (1983). A law of change in language. In B. Brainerd (Ed.), *Historical linguistics* (104-115). Bochum: Brockmeyer.
- Bailey, C.J. (1973). *Variation and linguistic theory*. Washington, DC: Centre for Applied Linguistics.
- Bauer, L. (1993). *Manual of information to accompany the Wellington Corpus of Written New Zealand English*. Wellington: Department of Linguistics, Victoria University of Wellington.
- Bauer, L. (1994). *Watching English change*. Harlow: Longman.
- Braun, A. (1982). *Studien zur Syntax und Morphologie der Steigerungsformen im Englischen*. Bern: Francke.
- Cheshire, J. (1999). Taming the vernacular: Some repercussions for the study of syntactic variation and spoken grammar. *Cuadernos de Filología Inglesa*, 8, 59-80.
- D’Arcy, A. & Tagliamonte, S.A. (2010). The prestige legacy of relative *who*. *Language in Society*, 39(3), 389-410.

- Denison, D. (2003). Log(ist)ic and simplistic S-curves. In R. Hickey (Ed.), *Motives for language change* (54-70). Cambridge: Cambridge University Press.
- Gordon, E., Campbell, L., Hay, J., MacLagan, M., Sudbury, A. & Trudgill, P. (2004). *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.
- Gordon, E., Hay, J. & MacLagan, M. (2007). The ONZE corpus. In J.C. Beal, K.P. Corrigan & H.L. Moisl (Eds.) *Creating and digitizing language corpora*, Vol. 2 (82-104). New York: Palgrave Macmillan.
- Guy, G. (1988). Advanced Varbrul analysis. In K. Ferrara, B. Brown, K. Walters & J. Baugh (Eds.), *Linguistic change and contact* (124-136). Austin: University of Texas at Austin.
- Hilpert, M. (2008). The comparative: Language structure and language use. *English Language and Linguistics*, 12(3), 395-417.
- Kroch, A.S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1(1), 199-244.
- Kytö, M. (1996). 'The best and most excellentest way': The rivalling forms of adjective comparison in Late Middle and Early Modern English. In J. Svartvik (Ed.), *Words: Proceedings of an international symposium* (123-144). Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien.
- Kytö, M. & Romaine, S. (1997). Competing forms of adjective comparison in Modern English: What could be *more quicker* and *easier* and *more effective*? In T. Nevalainen & L. Kahlas-Tarkka (Eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen* (329-352). Helsinki: Société Néophilologique.
- Kytö, M. & Romaine, S. (2000). Adjective comparison and standardisation processes in American and British English from 1620 to the present. In L. Wright (Ed.), *The development of Standard English, 1300-1800: Theories, descriptions, conflicts* (171-194). Cambridge: Cambridge University Press.
- Kytö, M. & Romaine, S. (2006). Adjective comparison in nineteenth-century English. In M. Kytö, M. Rydén & E. Smitterberg (Eds.), *Nineteenth-century English: Stability and change* (194-214). Cambridge: Cambridge University Press.
- Labov, W. (1966). *The Social Stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1994). *Principles of linguistic change*, Vol.1, *Internal factors*. Cambridge and Oxford: Blackwell.
- Leech, G. & Culpepper, J. (1997). The comparison of adjectives in recent British English. In T. Nevalainen & L. Kahlas-Tarkka (Eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen* (353-373). Helsinki: Société Néophilologique.
- Lindquist, H. (1998). The comparison of English disyllabic adjectives in -y and -ly in Present-day British and American English. In H. Lindquist, S.

- Klintborg, M. Levin & M. Estling (Eds.), *The major varieties of English. Papers from MAVEN 97* (205-212). Växjö: Acta Wexionensia.
- Lindquist, H. (2000). *Livelier or more lively?* Syntactic and contextual factors influencing the comparison of disyllabic adjectives. In J.M. Kirk (Ed.), *Corpora galore: Analyses and techniques in describing English* (125-132). Amsterdam: Rodopi.
- McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.
- Mitchell, B. (1985). *Old English syntax*, Vol. 1. Oxford: Clarendon Press.
- Mondorf, B. (2003). Support for *more-support*. In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of grammatical variation in English* (252-304). Berlin: Mouton de Gruyter.
- Mondorf, B. (2009). *More support for more-support. The role of processing constraints on the choice between synthetic and analytic comparative forms*. Amsterdam: John Benjamins.
- Mustanoja, T.F. (1960). *A Middle English syntax*, Part I: *Parts of speech*. Helsinki: Société Néophilologique.
- Pound, L. (1901). *The comparison of adjectives in the 15th and 16th centuries*. Heidelberg: Carl Winter.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. New York: Longman.
- Romaine, S. (1982). *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press.
- Schlüter, J. (2001). Why worser is better: The double comparative in 16th- to 17th-century English. *Language Variation and Change*, 13(2), 193-208.
- Scrivner, O. (2010). The white is more firm, and the yolk is rounder: Comparative variation in American English. Paper presented at NWAV 39, University of Texas at San Antonio, 5 November.
- Trudgill, P. (2004). *New dialect formation: On the inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.
- Weinreich, U., Labov, W. & Herzog, M.I. (1968). Empirical foundations for a theory of language change. In W.P. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics. A symposium* (95-195). Austin: University of Texas Press.