**Language-internal and -external factors during different stages of change**

**A corpus-based analysis of adjective amplification in Hong Kong, Indian, and Philippine English**

Martin Schweinberger, The University of Queensland, Australia

**Abstract**

This corpus-based study examines potential language variation and change in the adjective amplifier systems of Hong Kong, Indian, and Philippine English based on the International Corpus of English. The aim of the study is to ascertain if these Asian English varieties align with the global trend according to which innovative variants (*really* or *so*) are replacing *very* as the dominant adjective amplifier in informal spoken English. Furthermore, the study analyses which factors govern the use of adjective amplifiers during different stages of change. The results show that the adjective amplifier systems of Hong Kong and Indian English are remarkably stable across apparent-time while Philippine English shows notable signs of ongoing change. Conditional inference trees indicate that during stability and initial stages of change, as represented by the amplifier systems of Hong Kong and Indian English, intra-linguistic variables figure most prominently as factors that impact amplifier choice while extra-linguistic, social variables gain importance once a change is nearing mid-range, as seen in Philippine English. The results are interpreted to substantiate that language change is motivated by linguistic constraints but require social meaning to accelerate.

**Keywords**

Intensification, adjective amplifiers, language variation and change, conditional inference tree, very, really, so, Hong Kong English, Indian English, Philippine English

**Introduction**

This study investigates the adjective amplifier systems of Hong Kong English (HKE), Indian English (IndE), and Philippine English (PhiE) the effects of language-internal and language-external factors during ongoing change and evaluates if the importance of these factors differs across stages of change. To address this issue, the present study focuses on ongoing change in the adjective amplifier systems of Hong Kong, Indian, and Philippine English based on data from the International Corpus of English. As adjective amplification has been found to be a site of “fevered invention and competition that would be hard to come by elsewhere” (Bolinger 1972: 18), this domain lends itself to fine-grained analyses of change.

While there exists a substantial amount of research on research on the factors that impact language change (Labov 1994, 2001, 2010) and changes in adjective amplification (e.g. D’Arcy 2015, Tagliamonte 2008, Tagliamonte & Denis 2014), issues relating to when which factors are more or less impactful during changes in this domain remain unresolved. A key factor for why such issues were difficult to address is methodological: whereas most analyses of ongoing change use regression modelling to evaluate the impact of different factors during language change, tree-based models (see Tagliamonte & Baayen 2012) they are more appropriate to detect differences in the impact of factors across stages of change. Therefore, the present study uses Conditional Inference Trees to determine which factors are dominating during which phase of change.

One of the most consistent findings in previous research on changes in adjective amplification in informal spoken discourse has been the replacement of very by really (D’Arcy 2015, Tagliamonte 2008). The present study investigates if this trend holds true for Hong Kong, Indian, and Philippine English and if the underlying factors that drive this change vary across different stages of change in these Asian English varieties.

The present analysis shows that *very* is being replaced by *really* in Philippine English which is similar to the findings from research on change in adjective amplification in inner circle varieties. As such, the change that can be observed in Philippine English represents a vigorous to mid-range stage of change while the change is only in its incipient stage in Indian English. In Hong Kong English, the amplifier system represents a pre-change situation where the amplifier system is remarkably stable. The statistical analysis reveals that during stasis and incipient stages of change, amplifier sue is determined by language internal factors while extra-linguistic factors, in particular social factor, drive change once the change is accelerating and has become more vigorous.

In review of the converging results from previous research and the present study of Asian varieties of English, the paper adds credence to the hypothesis that linguistic change represents an “opportunistic process that reinforces social distinctions by associating them with particular linguistic variants” (Labov 2002). Applied to the present case this means that change is initially dominated by intra-linguistic factors and accelerates once the innovative variant is sufficiently associated with a social group that then serves as the extra-linguistic driving force during vigorous and mid-range change.

**Previous Research**

RQ1: Are Asian Englishes in-line with the trend among traditional L1 and other post-colonial varieties of English in that *very* is being replaced by *really*?

We hypothesise that HKE, IndE, PhiE align with other varieties: *really* is replacing *very* as the default adjective amplifier

In addition, we assume that PhiE which is more closely aligned with AmE and AusE will exhibit a more advanced stage of change while HKE and IndE, which are more aligned with GBE, will be lagging behind due to the more conservative nature of standard BrE with respect to adjective amplification.

**Data and Methodology**

The analysis uses data from the Hong Kong, Indian, and Philippine components of the International Corpus of English (ICE). While the ICE data were collected between 1990 and 2005 and the ICE data is thus not fully up-to-date, they have a shared design which allow us to draw meaningful comparisons between varieties of English. A further advantage of using ICE data stem from the fact that the corpus data is accompanied by socio-demographic information about the speakers which render the ICE components a very useful and interesting resource for variationist analyses. All ICE components consist of one million words (600,000 spoken and 400,000 written) from diverse spoken and written text types. For the current study, only data representing transcribed informal private spoken conversations were considered. This conversational section comprises roughly 200,000 words within each of the components.

The data processing, visualization, and analysis was performed in R (R Core Team 2020). During data processing, the data from the three ICE components was collapsed while retaining information of the origin of each data point. Collapsing the data guaranteed that the same processing steps were applied to the data irrespective of origin while retaining the ability to split the data at a later stage to enable comparisons between the three regional components. The spoken private dialogue section of each component was selected and part–of-speech tagged using the OpenNLP package (Hornik 2016).

Next, all adjectives (pos–tag JJ) were retrieved and it was determined whether adjective were preceded by an amplifier. Identifying amplifiers was based on a list of predefined amplifiers[[1]](#footnote-1) (see footnote XXX) and additional manual inspection. In addition, the sociodemographic information for each speaker that had uttered an adjective was added. This information consisted of the age (16-25, 26-41, 42+), gender (female versus male), and ethnicity of speakers (Chinese versus non-Chinese) as well as the date of the recording (before and including 1998 versus after 1998).

Furthermore, a sentiment analysis was applied to all adjective types (Jockers 2017). The sentiment analysis used in the current study is based on the Word-Emotion Association Lexicon (Mohammad & Turney 2013; cf. http://www.purl.org/net/NRCemotionlexicon) which comprises 10,170 terms and in which lexical elements are assigned scores based on ratings gathered through the crowd-sourced Amazon Mechanical Turk service. Raters were asked whether, according to their judgement, a given word was associated with one of eight basic emotions (ANGER, ANTICIPATION, DISGUST, FEAR, JOY, SADNESS, SURPRISE, TRUST; see Plutchik 1980, 1994). The resulting associations between terms and emotions are based on 38,726 ratings from 2,216 raters who answered a sequence of questions for each word which were then fed into the emotion association rating (cf. Mohammad & Turney 2013). Each term was rated 5 times. For 85 percent of words, at least 4 raters provided identical ratings. For instance, the words dark or tragic are more readily associated with SADNESS while words such as happy or beautiful are indicative of JOY and words like cruel or outraged may indicate ANGER. When applying the sentiment analysis to the present data, an adjective receives a negative-polarity score for every word that is associated with ANGER, DISGUST, FEAR, or SADNESS. For every word associated with ANTICIPATION, JOY, SURPRISE, or TRUST the tweet receives a score for positive polarity. This way, the sentiment analysis allows to test if the emotionality (positive, neutral, or negative) is correlated with changes in the use of specific amplifier variants.

The next step consisted in determining if the same amplifier type had occurred within a span of three adjective slots previously which is necessary to test for potential priming effects Szmrecsanyi 2005, 2006; Tulving & Schacter 1990: 301). Also, the token frequency of adjective type by corpus and age group was calculated to prevent frequency changes in adjective types to confound the analysis (see also Tagliamonte and Roberts 2005).

Next, misspellings of adjectives were corrected using the qdap package (Rinker 2020). Further, elements that were erroneously pos-tagged as adjectives, misclassified items, negated adjectives, and adjectives preceded by downtoners, adjectives that were not amplified by at least two different amplifier variants (to remove lexicalized expressions such as *right honorable*), forms that behave a-prototypical (e.g. *much*, *many*), as well as comparative and superlative forms were removed from the data.

Then, the syntactic context of each adjective was determined (attributive *the nice man* versuspredicative *The man is nice.*). Also, each adjective was assigned a gradability score. This score represents the likelihood of an adjective type to occur in comparative contexts against the likelihood of its occurrence in a non-comparative context. As such, the gradability score reflects the tendency of adjective to be more or less gradable. Also, each adjective was categorized semantically based on Dixon’s (1977) basic semantic adjective types (cf. also D’Arcy (2015); Tagliamonte and Roberts (2005); Tagliamonte (2006, 2008)). However, after testing the usefulness and validity of this fine-grained classification using conditional inference trees, the categories proposed by Dixon (1977) were simplified to only differentiate between adjective that denote value-related meanings (e.g. *good* or *bad*) and all other adjectives. Furthermore, due to the overall sparsity of variants other than *very*, *really*, and *so*, all other variants were collapsed into a bin category *other*. Similarly, all adjective types other than *different* were collapsed into a bin category as the multivariate analysis revealed that *different* was the only adjective type that significantly affected amplifier choice. The final step of data processing consisted in a manual cross–evaluation of the data. The final data is summarized in Table XXX.

A summary overview of the variables analysed in the present study is provided in Table XXX.

**Statistical analysis**

The statistical analysis of the data uses frequency tabulation to tap into apparent-time trajectories of amplifier use and conditional inference trees (see Hothorn et al. 2006; Hothorn, Hornik & Zeileis 2006) based on the partykit package (Hothorn & Zeileis 2015) to determine which factors correlate with the use of different variants. Conditional inference trees (CIT) iteratively partition the data so as to minimize residual deviance while maximizing the distinctness of the resulting groups. Appropriate alternatives such as multinomial mixed-effects regression models could also be used but their output is less readily interpretable.

CITs were applied to the combined data as well as to each variety separately. This allowed us to determine which factors correlated with amplifier choice overall but also within each variety.

**Results**

The following presents the results of the multivariate analysis which used conditional inference trees to ascertain which factors impact amplifier choice across and with HKE, IndE, and PhiE. We begin with the depiction of the apparent-time trajectory of amplifier types by syntactic context (see Figure 1).

|  |
| --- |
|  |

Figure 1 shows that the amplifier systems in the Hong Kong and Indian ICE data are remarkably stable with very being consistently the dominant adjective amplifier across age groups and syntactic contexts. A minor difference between HKE and IndE is visible for the use of so in predicative contexts: while *so* has decreased across apparent-time in HKE, it has slightly increased across apparent time in IndE. In contrast, PhiE shows signs of lexical replacement that mirrors trends in other varieties of English with *very* decreasing and *really* increasing across apparent time in both syntactic contexts. In addition, so has already replaced very as the dominant amplifier variant in predicative contexts. However, *so* remains to be constrained from entering attributive contexts and its use is thus confined to predicative contexts.

We now turn to the results of the CITs. The first CIT, which was applied to the data set which contained data points from all three regional varieties, confirms that the regional varieties differ remarkably with a separation between HKE and IndE on the one hand and PhiE on the other representing the first, and thus most important split. Overall, intra-linguistic factors appear to figure more prominently in IndE and HKE, while social variables dominate amplifier choice in PhiE.

|  |
| --- |
| Figure 1: Conditional Inference Tree for HKE, IndE, and PhiE. |

Figure 1 shows that in HKE and IndE adjective amplification is governed by intra-linguistic factors, only the gender of speakers has an effect and only within a small subset of the Indian ICE data. In contrast, the syntactic context of adjectives in the only intra-linguistic factor that impacts amplifier use in PhiE, all other partitions are based on the age and gender of speakers. To ascertain this impression, we now inspect the CIT which were applied to each variety separately and begin with inspecting the CIT that was fit to the Hong Kong ICE data.

|  |
| --- |
| Figure 2: Conditional Inference tree for the ICE Hong Kong data. |

Figure 2 confirms that adjective amplification in HKE is dominated by *very* and governed only by language internal factors with the syntactic context being the most distinctive variable. In attributive context, very is the uncontested default amplifier with substantially smaller shares of adjectives being amplified by *really* and *other* variants. In predicative contexts, *different* is over-proportionately used with other amplifiers (predominantly *completely*). If adjectives in predicative contexts represent value semantics (*good*, *bad*, etc.) speakers show an elevated likelihood to select *really* and, to a lesser degree, *so* as alternatives to *very*. If the adjectives in predicative contexts do not represent value semantics, speakers tend to select *so* as an alternative to *very* while disfavouring *really*.

We now turn to the CIT that was fit to the Indian ICE data.

|  |
| --- |
| Figure 3: Conditional Inference tree for the ICE India data. |

Figure 3 shows that *different* is amplified predominantly by *other* amplifiers (which in this case is caused by an over-proportional bigram frequency of *completely different*). Across all other leaves *very* is the dominant amplifier in the Indian ICE data, however, *other* amplifiers also figure prominently for predicative value adjectives uttered by older speakers. In contrast, *so* is over-proportionately used with predicative adjectives particularly if these adjectives do not represent value semantics (*good* or *bad*) and particularly among female speakers.

We now turn to the CIT that was fit to the Philippine ICE data.

|  |
| --- |
| Figure 4: Conditional Inference tree for the ICE Philippine data. |

Figure 4 differs substantially form the CIT results of IndE and HKE.

**Discussion**

Summary

The analysis shows that . . .

- the trajectory of change in the adj. amplifier system of PhiE is approximating the trajectories of other varieties (NZE, IrE, AusE, etc.)(see e.g. D’Arcy 2015)

! only PhiE confirms H11

- change in PhiE is substantially further advanced compared to HKE and IndE ! confirmation of H12

- during stasis and initial stages of change, intra-linguistic factors dominate

- once change is accelerating, extra-linguistic (social) factors become more important (see Labov 2002)

Discussion

- During initial stages of change, the type of the amplified adjective and its linguistic properties determine amplifier use

- Once an amplifier variant attains social meaning, it will (further) increase in use: linguistic change piggybacks on social stratification

“Sociolinguistic variation is parasitic upon such linguistic variation. It is an opportunistic process that reinforces social distinctions by associating them with particular linguistic variants. [. . . ] The use of linguistic forms to increase distinctiveness of particular groups is a driving force for the acceleration of change” (Labov 2002)

Outlook

Could this be a universal mechanism/pattern?

Test if the mechanisms. . .

- can be shown to have worked in analogous changes in

English (e.g. 3rd p. sg. ind. morpheme: <eth> ! <(e)s> )

- can be shown to have worked in analogous changes in languages other than English

**References**

Aijmer, K. (2011). Are you totally spy? a new intensifier in present-day american english. In S. Hancil (Ed.), Marqueurs discursifs et subjectivité, pp. 155–172. Rouen: Universités de Rouen and Havre.

Aijmer, K. (2018). That’s well bad. some new intensifiers in spoken in british english. In V. Brezina, R. Love, and K. Aijmer (Eds.), Corpus Approaches to Contemporary British English, pp. 60–95. New York and London: Routledge.

Barnfield, K. and I. Buchstaller (2010). Intensifiers on tyneside - longitudinal developments and new trends. English World-Wide 31(3), 252–287.

Bauer, L. and W. Bauer (2002). Adjective boosters in the english of young new zealanders. Journal of English Linguistics 30, 244–257.

Bolinger, Dwight. 1972. *Degree words*. The Hague: Mouton.

Brinton, L. J. and L. K. Arnovick (2006). The English Language: A Linguistic History. Oxford: Oxford University Press.

D’Arcy, Alexandra F. 2015. Stability, stasis and change - the longue duree of intensification. *Diachronica* 32(4): 449–493.

Dixon, R. M. W. (1977). Where have all the adjectives gone? Studies in Language 1, 19–80.

Fuchs, R. (2016). Register variation in intensifier usage across asian englishes. In H. Pichler (Ed.), Discourse-Pragmatic Variation and Change: Insights from English, pp. 185–213. Cambridge: Cambridge University Press.

Fuchs, R. (2017). Do women (still) use more intensifiers than men? International Journal of Corpus Linguistics 22(3), 345–374.

Hothorn T, Hornik K, Van de Wiel MA, Zeileis A (2006). A Lego System for Conditional Inference. The American Statistician, **60**(3), 257–263.

Hothorn T, Hornik K, Zeileis A (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics, **15**(3), 651–674.

Hothorn T, Zeileis A (2015). partykit: A Modular Toolkit for Recursive Partytioning in R. Journal of Machine Learning Research, **16**, 3905–3909.

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

Jockers, M. L. (2017). Syuzhet 1.0.4 now on cran. http://www.matthewjockers.net/2017/12/16/syuzhet-1-0-4/.

Labov, W. (1966). The Social Stratification of English in New York City. Washington DC: Center for Applied Linguistics.

Labov, W. (1972). Sociolinguistic patterns. Philadelphia, PA: University of Pennsylvania Press.

Labov, W. (1994). Principles of Language Change: Internal Factors, Volume 1. Oxford: Blackwell.

Labov, William. 1994*. Principles of Linguistic Change. Vol. 1, Internal Factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of Linguistic Change. Vol. 2, Social Factors*. Oxford: Blackwell.

Labov, William. 2002. Driving forces in linguistic change. Paper presented at *the International Conference on Korean Linguistics*.

Labov, William. 2010. *Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors*. Oxford: Blackwell.

Macaulay, R. (2006). Pure grammaticalization: The development of a teenage intensifier. Language Variation and Change 18, 267–283.

Núñez Pertejo, P. and I. Palacios (2014). That’s absolutely crap, totally rubbish. the use of intensifiers absolutely and totally in the spoken language of british adults and teenagers. Functions of Language 21(2), 210–237.

Palacios, I. and P. Núñez Pertejo (2012). He’s absolutely massive. it’s a super day. madonna, she is a wicked singer. youth language and intensification: A corpus-based study. Text and Talk 32(6), 773–796.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985). A Comprehensive Grammar of the English Language. London & New York: Longman.

Schweinberger, M. (2017). Using intensifier-adjective bi-grams to investigate mechanisms of change. Paper presented at ICAME38. Prague, 27/5/2017.

Tagliamonte, S. (2006). ”so cool, right?”: Canadian english entering the 21st century. The Canadian Journal of Linguistics/La revue canadienne de linguistique 51(2), 309–331.

Tagliamonte, Sali. 2008. So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics* 12(2): 361–394.

Tagliamonte, Sali & Harald R. Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24: 135–178.

Tagliamonte, Sali. A. & Derek Denis. 2014. Expanding the transmission/diffusion dichotomy: Evidence from Canada. *Language* 90(1): 90–136.

Tagliamonte, S. and C. Roberts (2005). So weird; so cool; so innovative: The use of intensifiers in the television series friends. American Speech 80(3), 280–300.

Tagliamonte, S. A. (2011). Variationist sociolinguistics: change, observation, interpretation, Volume 39. Malden & Oxford: John Wiley & Sons.

Wagner, S. (2017a). Amplifier-adjective 2-grams world-wide: focus on pretty. Paper presneted at ICAME 37. Charles University Prague, 27/5/2017.

Wagner, S. (2017b). Totally new and pretty awesome: Amplifier–adjective bigrams in glowbe. Lingua 200, 63–83.

1. absolutely, actually, aggressively, amazingly, appallingly, awfully, badly, bloody, certainly, clearly, dead, completely, considerably, crazy, decidedly, definitely, distinctly, dreadfully, enormously, entirely, especially, exactly, exceedingly, exceptionally, excruciatingly, extraordinarily, extremely, fiercely, firmly, frightfully, fucking, fully, genuinely, greatly, grossly, heavily, highly, hopelessly, horrendously, hugely, immediately, immensely, incredibly, infinitely, intensely, irrevocably, mad, mega, mighty, most, much, obviously, openly, overwhelmingly, particularly, perfectly, plenty, positively, pretty, profoundly, purely, real, really, remarkably, seriously, shocking, significantly, so, specially, specifically, strikingly, strongly, substantially, super, surely, terribly, terrifically, total, totally, traditionally, true, truly, ultra, utterly, very, viciously, wholly, wicked, wildly [↑](#footnote-ref-1)