From *insanely jealous* to *insanely delicious*: Computational models for the semantic bleaching of English intensifiers

Yiwei Luo[1] **Dan Jurafsky**[1,2] **Beth Levin**[1]
[1]Stanford Linguistics [2]Stanford Computer Science
{yiweil, jurafsky, bclevin}@stanford.edu

Abstract

We introduce novel computational models for modeling semantic bleaching, a widespread category of change in which words become more abstract or lose elements of meaning, like the development of *arrive* from its earlier meaning 'become at shore.' We validate our methods on a widespread case of bleaching in English: de-adjectival adverbs that originate as manner adverbs (as in *awfully behaved*) and later become intensifying adverbs (as in *awfully nice*). Our methods formally quantify three reflexes of bleaching: decreasing similarity to the source meaning (e.g., *awful*), increasing similarity to a fully bleached prototype (e.g., *very*), and increasing productivity (e.g., the breadth of adjectives that an adverb modifies). We also test a new causal model and find evidence that bleaching is initially triggered in contexts such as *conspicuously evident* and *insanely jealous*, where an adverb premodifies a semantically similar adjective. These contexts provide a form of "bridging context" (Evans and Wilkins, 2000) that allow a manner adverb to be reinterpreted as an intensifying adverb similar to *very*.

1 Introduction

Developments in computational semantics and availability of large diachronic corpora have renewed interest in studying historical semantic change. Recent work has moved away from documenting and qualitatively categorizing types of changes (Bréal, 1964; Stern, 1931) to focus on detecting semantic shifts (Gulordava and Baroni, 2011; Rosenfeld and Erk, 2018; Frermann and Lapata, 2016; Mitra et al., 2014; Kulkarni et al., 2015), distinguishing gradual linguistic drifts from cultural ones (Hamilton et al., 2016a) and assessing laws of change (Hamilton et al., 2016b; Dubossarsky et al., 2017; Xu and Kemp, 2015; Ramiro et al., 2018; Luo and Xu, 2018).

Building off prior work, we propose the first computational study of semantic bleaching, one of the most widespread changes in word meaning. Work in historical linguistics characterizes bleaching as an abstraction or loss of some initial elements of meaning, such as in the example *arrive*, which has broadened from 'become at shore', or *amazing*, which has undergone a change from 'stupefying' to 'great'. However, we know very little about how this change happens as a quantifiable and continuous process. For example, can we measure to what extent a bleached word continues to bear its root meaning? How much of the meaning of "awefulness" does *awfully* have, and to what extent does *awfully* now mean *very*? Finally, the fundamental question of what drives bleaching remains open.

Answering these questions requires a way to model the nuances of semantic bleaching separately from general semantic shifts. Thus, our work asks the following:

Q1: Can we build computational models of the bleaching process that match known semantic reflexes of bleaching?

To answer this question, we develop methods for quantifying three known reflexes of bleaching from the theoretical literature on semantic change: loss of original lexical meaning, gain of bleached target meaning, and increasing productivity. We focus on the case of English de-adjectival adverbs (***awfully*** *nice,* ***insanely*** *delicious*), which originally have a manner meaning derived from the semantics of their root adjective and later bleach into intensifying adverbs (or *intensifiers*) (Tab. 1). We choose this case of bleaching as it represents an open class of semantically diverse adverbs that experience exceptionally rapid change and speaker innovation (Bolinger, 1972; Peters, 1994).[1].

[1]Though he focuses on synchronic properties of degree words, Bolinger (1972, 18) observes: "[Intensifiers] afford

Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, pages 1–13
Florence, Italy, August 2, 2019.

Original usage	Bleached usage
awfully behaved	**awfully** nice
wildly flailing	**wildly** easy
insanely muttering	**insanely** delicious
abundantly endow	**abundantly** at ease
singing **terribly**	**terribly** sorry
aggressively demanded	**aggressively** sunny

Table 1. Examples of the bleaching phenomenon: de-adjectival **adverbs** in their original, manner usage and in their bleached, intensifier usage.

Next, we apply our methodology for modeling bleaching to answer open questions concerning *how* bleaching happens over time:

Q2: Can bleaching be explained in terms of reanalysis, by which certain contextual factors lead to one interpretation being favored over another?

Q3: If bleaching is a form of reanalysis, what are the contexts that trigger this re-interpretation?

We use the same semantically diverse set of bleaching de-adjectival adverbs to formulate and test hypotheses pertaining to these questions (Study 2, Sec. 4), building on previous diachronic work on intensifiers that have focused on a single word (Lorenz (2002), Macaulay (2006), Beltrama and Bochnak (2015), Tagliamonte (2008)). In particular, we hypothesize that a high semantic similarity between an adverb and the adjectives that it initially modifies is a crucial contextual factor that triggers the reanalysis of manner adverbs into intensifiers. This criterion (exemplified by collocations such as *conspicuously evident*, *terribly gruesome*) is what allows a manner adverb to be interpreted as an intensifier in the first place.

2 Methods for modeling bleaching

We translate three known reflexes of semantic bleaching from the literature—loss of lexical meaning; gain of intensifier meaning; increased productivity—into relationships between word embeddings and n-gram parse context. For our n-gram data, we use the English fiction portion of the Google Books English n-grams corpus (Lin et al., 2012) and for the historical word embeddings, we use the HistWords dataset trained on the same portion of the n-gram dataset (Hamilton et al., 2016b). The full corpus spans the years 1800 to 1999 but we restrict our temporal range to 1850 to 1999, inclusive, due to data sparsity. We test two different sets of HistWords embeddings: Word2Vec (W2V) representations and SVD representations. All data are aggregated to the granularity of decades, yielding 15 decades total.

a picture of fevered invention and competition it would be hard to come by elsewhere [...] They are the chief means of emphasis for speakers for whom all means of emphasis quickly become stale and need to be replaced."

2.1 Similarity of adverbs to *very* (SIMVERY)

As a manner adverb bleaches into an intensifying adverb, we expect the meaning of the adverb to grow more similar to the meaning of *very*, the prototypical example of a completely bleached intensifier (Peters, 1994). We measure this similarity via the cosine similarity between the HistWords embedding for an adverb and the embedding for *very*, both retrieved for a given decade. The bleached status of *very* is empirically verified in the embedding space: the self-similarity between consecutive decades is comparable to words expected to change extremely little over time, such as determiners, numerals, and pronouns (*the, two, three, four, them, they, us*, etc.) (Pagel et al., 2007).

2.2 Similarity of adverb to original lexical meaning (SIMLEX)

As a manner adverb like *awfully* bleaches into an intensifier, its meaning diverges from its root adjective's lexical meaning of "awfulness." We formalize this intuition of a bleaching adverb's divergence from its lexical meaning as the average cosine similarity of an adverb to a set of lemmas (L) related to its lexical meaning (eq. 1). We constructed these lemma sets by retrieving WordNet (Miller, 1998) synonyms for the root adjective and supplementing these with additional synonyms according to the *Oxford English Dictionary (OED)* (Simpson et al.) (Tab. 2).

Adverb	Lexical source lemmas
disgustingly	filthy, filth, repulsive, aversion
beautifully	elegance, elegant, style, gorgeous, beauteous
wildly	savage, rage, fierce, barbarian, uncivilized
remarkably	impact, stun, awe, wonder, amazement, terror

Table 2. Examples of adverbs and lemmas related to the lexical source meaning for computing `SimLex`.

$$\textsc{SimLex}(adv) = \frac{1}{|L|} \sum_{l_k \in L} \texttt{sim}_t(adv, l_k), \quad (1)$$

where L is a lemma set of lexical meanings and $\texttt{sim}_t(adv, l_k)$ is the cosine similarity at time t between an adverb and a lemma l_k in L.[2]

2.3 Productivity of adverb (BREADTH)

As an adverb bleaches, we expect to see greater productivity, i.e., an increase in the variety of the adjectives that it modifies. For example, we expect to see *terribly* modifying a greater range of sentiment adjectives over time. We suggest two distinct ways to quantify this semantic breadth. The first is *type diversity* (TYPEDIV)—the number of types modified—which is shown in Bybee (1995) to be important in determining productivity. The second is BREADTH, which we measure as negative cosine similarity of the adjectives, to capture how semantically similar the set of modified adjectives is. The more similar the adjectives modified by an adverb are to each other, the less semantically broad they are. This more general approach is useful since an adverb might modify a larger number of distinct adjectives while becoming more restricted in the meanings of adjectives that it modifies.

We extract all adjectives modified by an adverb for a given decade from the Google Syntactic n-grams corpus (Goldberg and Orwant, 2013). To calculate a single value for similarity among many adjectives, we subset the top 50 adjectives ranked by log odds, then take the grand average of all the pairwise similarities between each distinct adjective type (eq. 2). We also weight each pairwise similarity by each adjective's odds of being modified. The BREADTH B of an intensifier I at time t can be expressed as:

$$B(I, t) = -\sum_{a_i \in A_{I,t}} \sum_{\substack{a_j \in A_{I,t} \\ i \neq j}} \texttt{sim}(a_i, a_j) o(a_i) o(a_j) \quad (2)$$

where $A_{I,t}$ is the set of all adjectives modified by an intensifier I at time t, $\texttt{sim}(\cdot,\cdot)$ is the cosine similarity between two words, and $o(\cdot)$ is the odds of an adjective being modified by an adverb.

3 Study 1: Do our methods capture bleaching?

We hypothesize that our methods can be used to distinguish adverbs undergoing bleaching into intensifying adverbs from non-bleaching control adverbs. In particular, we expect to see significant correlations among the set of intensifiers between each metric and time in the following directions (Tab. 3) after fitting linear regressions on $\{y_t, t\}_{t=1850}^{2000}$, where y_t represents a bleaching metric evaluated at decade t.:

metric	sign of slope over time
SIMVERY	+
SIMLEX	–
BREADTH	+

Table 3. Predicted correlations between each bleaching metric over time (as the dependent variable) and time (as the independent variable) for bleaching adverbs.

To test these predictions, we introduce a set of bleaching intensifiers and a frequency-matched control set of non-bleaching adverbs. We expect to see significantly increasing similarity to *very* (SIMVERY), decreasing similarity to original meaning (SIMLEX), and increasing productivity (BREADTH) over time for intensifiers, and we expect that the slopes over time of these metrics are significantly greater for intensifiers than for the control adverbs.

3.1 Datasets

For both the intensifier and control sets, we restrict to de-adjectival adverbs (also known as *ly* type adverbs).[3] We sample these de-adjectival adverbs from lexical classes of adjective roots identified by Bolinger (1972) and supplement these with synonyms from WordNet (Miller, 1998). The result is a set of 250 intensifiers, shown partially in Table 4. (See Appendix A for the full set.)

Our control set consists of 178 frequency-matched adverbs sampled from the British National Corpus (BNC) (shown partially in Tab. 5, see Appendix B for the full set).[4] We obtained

[2]To increase the robustness of this metric, we restricted lemmas in L to those whose embeddings remained relatively stable over time by verifying that their self-similarities over successive decades did not differ significantly from a highly stable word set composed of determiners, numerals, and pronouns (t = 8.2e-01, p = 0.85).

[3]We also discard adverbs for specific years due to OOV-ness at random from either the W2V or SVD embeddings.

[4]Examples of usage taken from the British National Corpus (BNC) were obtained under the terms of the BNC End User Licence. Copyright in the individual texts cited resides with the original intellectual property right holders. For information and licensing conditions relating to the BNC, please see the web site at http://www.natcorp.ox.ac.uk/.

Root adjective type	Examples
magnitude	enormously, vastly, immensely, greatly, abundantly, massively
strength	overpoweringly, strongly, vigorously, exuberantly
singularity	distinctly, unusually, abnormally, mysteriously
evaluation	marvellously, brutally, dramatically, luxuriously, terribly, monstrously
irremediability	desperately, abominably, pathetically, disastrously
purity and veracity	undoubtedly, thoroughly, absolutely, fully, sincerely

Table 4. Examples of intensifiers, categorized by root adjective type according to Bolinger (1972).

average (relative) frequency estimates from the Google Books corpus over the period 1850-1990 and we selected the control adverbs from semantic categories such as time adverbs (*firstly, formerly, finally, temporarily, eventually*) and speed adverbs (*rapidly*, *quickly, slowly, promptly*), avoiding semantic categories of intensifiers that have been identified in the literature (Bolinger, 1972; Morzycki, 2008; Nouwen, 2011; Paradis, 1997).

abruptly	accordingly	frankly
ironically	locally	loudly
nationally	newly	officially
privately	quietly	simultaneously
happily	neatly	originally

Table 5. Examples of control adverbs.

3.2 Comparison of BREADTH to TYPEDIV

To determine whether or not BREADTH is independent from TYPEDIV (the number of adjective types modified by an adverb), we compute Spearman correlation coefficients between the metrics for individual adverbs as well as a single correlation between BREADTH and TYPEDIV averaged across all adverbs. We find that there are no significant correlations between average TYPEDIV and average BREADTH, nor do we find significant correlations between the two metrics within individual adverbs, indicating that our weighted BREADTH measure captures differences in productivity independent from the number of types that an adverb modifies. In fact, 200 of the 250 intensifiers in our dataset show a decrease in the number of types they modify within the last 5 decades of our data, but an increase in BREADTH.

3.3 Study 1 Results

We computed the 4 metrics (SIMVERY, SIMLEX, BREADTH, and TYPEDIV) on the intensifier and control adverbs described in Section 3.1 over the 14 decades from 1850 to 1990. As a reminder, SIMVERY measures an adverb's average semantic similarity to *very* and SIMLEX measures an adverb's average semantic similarity to its root adjective meaning (e.g., *completely* to {*full, entire, whole, ...*}). Both BREADTH and TYPEDIV measure the collocational freedom of an adverb, with the latter taking into account only the type diversity of adjectives that the adverb modifies and the former also incorporating the semantic similarity of those modified adjectives to each other. We then fit linear regressions with each bleaching metric as the dependent variable and time as the independent variable.[5] We take the natural log of BREADTH so that values are linear after weighting by adjective frequencies. We also compute each bleaching metric separately with Word2Vec (W2V) and SVD embeddings, expecting the strength and direction of the correlations to be unaffected by the choice of embedding.

The 10 most and least bleached intensifiers by each metric using W2V embeddings for 1990 are shown in Tab. 7; examples showing increasing BREADTH over the period 1850-1990 are shown in Tab. 6. A visual of increasing BREADTH is shown in Fig. 1.

The results of our regressions somewhat support our predicted temporal correlations (Fig. 2). As a caveat, we note that the increasing size of the syntactic n-grams corpus over time likely biases BREADTH, since a larger corpus has more contexts for each word, thus potentially inflating the strength of the correlation with time. While weighting BREADTH by each adjective's likelihood of being modified may mitigate this bias to an extent (since each likelihood is expected to decrease as corpus size increases), we recognize that future work should seek more robust forms of nor-

[5]We performed all regressions using ordinary least squares models in the StatsModels Python module (Seabold and Perktold, 2010).

	1850	1990
terribly	deficient, deformed, diseased, beaten, broken, fatal, unorthodox, guilty...	relieved, smitten, small, important, valid, goodlooking, generous, tired, pregnant...
abundantly	fat, large, flowing, fertile, rejoicing, grateful...	available, fraught, intelligible, loud, eager, familiar...
enormously	rich, large, high, long, great, fat, wealthy, thick...	popular, successful, important, complex, influential, difficult, helpful...

Table 6. Three bleaching adverbs and examples of adjectives they modify in the Google Books corpus at 1850 vs. 1990, showing an increase in productivity of the bleaching adverb.

	most bleached	least bleached
SimVery	**extremely**, terribly, truly, awfully, *definitely*, remarkably, **absolutely**, precisely, honestly, seriously	amply, vigorously, richly, *heavily*, violently, mysteriously, profusely, severely, furiously, miraculously
SimLex	entirely, decidedly, *heavily*, supremely, **particularly**, sorely, literally, deeply, especially, sharply	pleasantly, abundantly, enthusiastically, intensely, delightfully, *definitely*, furiously, curiously, *evidently*, **profusely**
Breadth	wholly, completely, **particularly**, deeply, *evidently*, distinctly, **absolutely**, **extremely**, perfectly, clearly	grievously, gorgeously, stupendously, surpassingly, outrageously, miraculously, deliciously, extravagantly, **profusely**, ludicrously

Table 7. The 10 most and least bleached intensifiers in 1990 according to each metric computed using W2V embeddings. Intensifiers **in bold** are most or least bleached according to more than one metric. Intensifiers *in italics* are categorized as most bleached by one metric but least bleached by another.

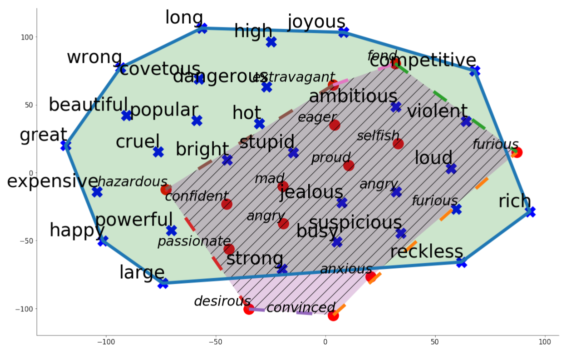

Figure 1. t-SNE visualization of adjectives modified by *insanely* in 1850 (plotted as circles; italicized) vs. in 1990 (plotted as x's), with convex hulls of each decade's adjectives shown in hatched purple and solid green, respectively, showing that the category of adjectives that are modified by *insanely* has expanded over 140 years.

malization.

The signs of the slopes match our predictions for all metrics and across embedding types for the intensifier set. Moreover, the strength of the correlation is significant for SimVery (p<1e-01) as well as for Breadth (p<1e-4) when computed using both W2V and SVD embeddings. For SimLex, the strength of this correlation is also significant (p<1e-05), but only when computed using W2V embeddings.

For the control set, we find that there are no significant correlations for SimLex computed using either embedding type (p>0.50), which matches our predictions. Nor do we find significant correlations for SimVery when computed using SVD embeddings. However, we do find a significant positive slope (p<1e-06) for SimVery+W2V, indicating that the control adverbs in our dataset are also becoming more similar to *very* over time. Nevertheless, the slope over time is still significantly greater for intensifiers than control adverbs (t = 3.1, p<1e-02).

Finally, the correlation for Breadth is significant for both intensifiers and control when computed using W2V embeddings as well as using SVD embeddings (p<1e-63, p<1e-05), suggesting that our current metric for change in productivity might be heavily dependent on corpus size. While we did not find any correlations between Breadth and TypeDiv, we find that the latter measure of productivity also shows significant trends of increase for both intensifiers and control (again, likely due to increasing corpus size). However, we find that the size of the slope for

TYPEDIV is significantly greater for intensifiers than control (t = 4.28, p<1e-04), indicating that this metric can identify a bleaching adverb given a control set of non-bleaching adverbs.

3.4 Discussion

We find that the combinations SIMVERY+SVD and SIMLEX+W2V successfully distinguish between bleaching and non-bleaching adverbs, yielding significant slopes over time for the former and no significant slopes for the latter. Surprisingly, SIMVERY+W2V shows a significant increase over time for both intensifiers and control, despite the fact that the principal meaning difference between the two sets is the new meaning of intensification that only the bleaching adverbs acquire. However, we note that this metric is still useful for identifying bleaching adverbs when a control set of non-bleaching adverbs is defined, since the size of the slope is significantly larger for the former. We find that BREADTH does not work in distinguishing bleaching from non-bleaching adverbs, most likely due to its dependence on corpus size, though possibly also because it captures changes that are not due strictly to bleaching (such as metaphorical extension, though we do not investigate this suspicion here). However, we find that TYPEDIV (just as SIMVERY+W2V) does work in the setting of a control set being available, as the size of the slope is significantly greater for intensifiers compared to control adverbs.

It is also possible that SIMLEX may show some bias toward adverbs that are less morphologically transparent with respect to their root—for example, we see that *sorely, especially,* and *decidedly* are among the 10 most bleached intensifiers identified by SIMLEX in Tab. 7. We hope to explore refinements to SIMVERY[6] and our two productivity measures (BREADTH and TYPEDIV) in future work that may better distinguish between bleaching and non-bleaching adverbs even without a control set readily available.

4 Study 2: Testing a causal theory

Ultimately, we are interested in modeling bleaching in order to test hypotheses concerning *how* a change like ***awfully*** *behaved* to ***awfully*** *nice* took place. In particular, we hypothesize a reanalysis-driven account of this change:

H1: When an adverb begins to modify adjectives that are semantically similar to itself, the adverb begins to be re-interpreted as an intensifier.

We now turn to the logic behind our hypothesis and the predictions made by our theory.

4.1 A theory of reanalysis-driven bleaching

For our causal theory, we adopt the framework of reanalysis as in work by Bybee et al. (1994), Hopper and Traugott (2003), and Evans and Wilkins (2000). In these works, interpretations that initially arise out of pragmatic enrichment become conventionalized over time due to regularly occurring contexts that provide support for the enriched interpretation. Following Evans and Wilkins (2000), we refer to these supporting contexts as "bridging contexts."

In the case of the reanalysis of a manner adverb into an intensifier, we hypothesize that the bridging context crucially involves the premodification of an adjective, *A*, that denotes a semantically similar property. To develop an intuition for how this criterion can give rise to the contextual ambiguity *very A*, we refer to examples (1-3) below from The Corpus of Historical American English (*COHA*) (Davies, 2010-). In (1-3)(b), the adverb and modified adjective denote independent properties: abnormalness is independent from being developed, awfulness is independent from being behaved, etc. However, in (1-3)(a), both adverb and adjective are associated with a shared semantic property such that the adverb reiterates the modified adjective in a way that is analogous to intensification.

(1) a. There is an **abnormally** disproportionate lack of demand.
  b. The most **abnormally** developed organs [...]

(2) a. [...] but it has left these rooms **awfully** dirty.
  b. [...] most **awfully** behaved girl she had ever met.

(3) a. The scenery on the river was **beautifully** picturesque [...]
  b. The country is **beautifully** broken, highly fertile, and cultivated like a garden.

Our theory hypothesizes that only for the (a) contexts involving an adverb and adjective pair both

[6]We perform the same analyses with a modified version of SIMVERY that measures the average cosine similarity of an adverb to {*very, really*} but find that the results are slightly poorer in distinguishing bleaching from non-bleaching.

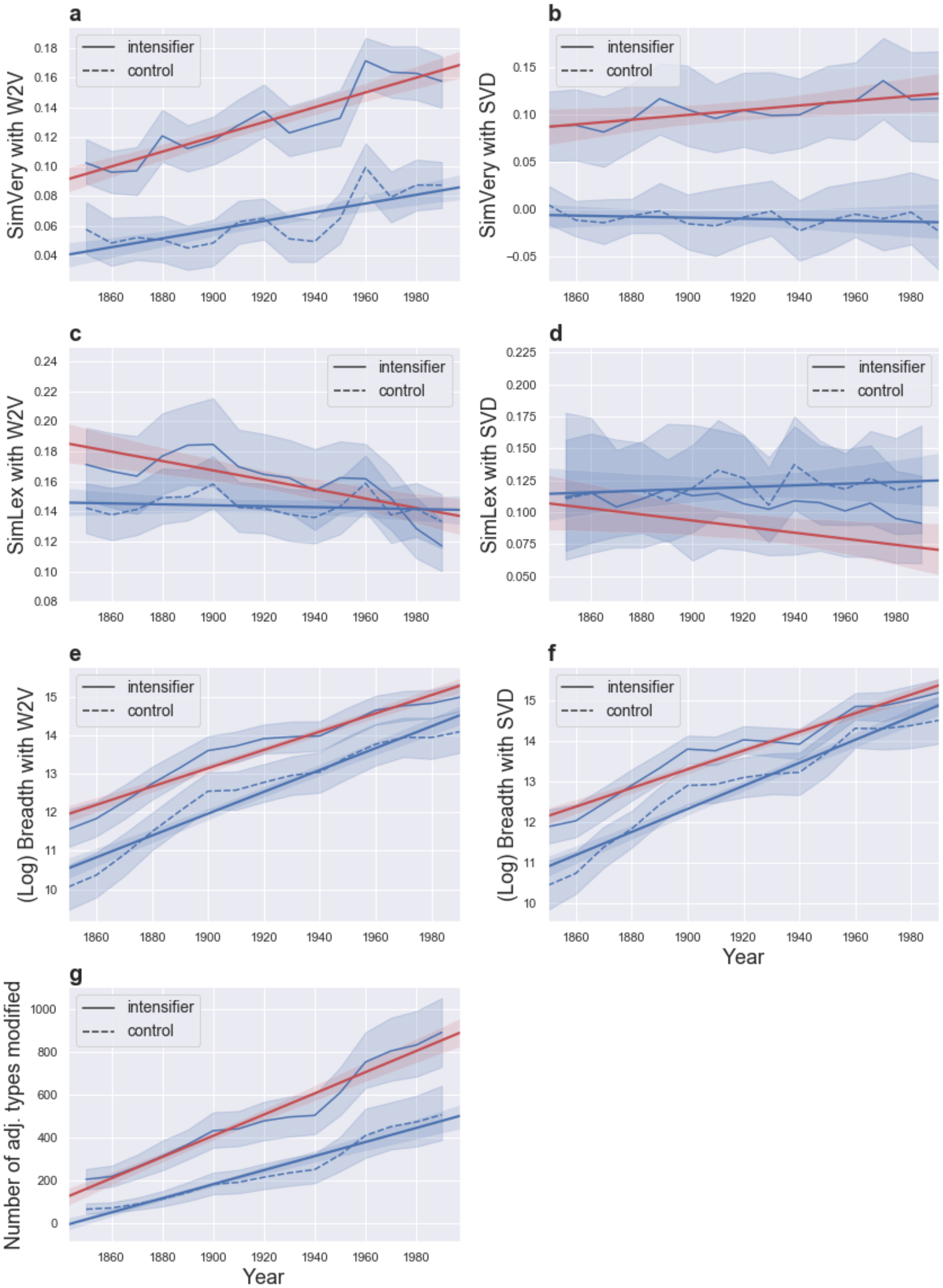

Figure 2. Raw extents of bleaching over time and lines of best fit from OLS linear regressions, showing partial confirmation of predicted trends. Intensifiers show significantly greater increases in W2V similarity to *very* (SIMVERY) over time compared to control adverbs (**a**), intensifiers show increasing SVD similarity to *very* over time while control adverbs show no trend (**b**). Intensifiers show decreasing W2V similarity to their original lexical meanings (SIMLEX) over time whereas control adverbs show no trend (**c**). Neither intensifiers nor control show a significant trend with SIMLEX using SVD embeddings (**d**). Intensifiers and control adverbs both show increasing productivity over time measured as BREADTH (**e-f**) and as raw type diversity, but intensifiers show significantly greater increases over time compared to control for TYPEDIV (**g**). Error bars on raw values show 95% bootstrap confidence intervals.

related to a single property p does their combination yield a synergy such that language users can infer the meaning 'very p.' As these bridging contexts increase in number, there is eventually enough evidence for users to infer the adverbial meaning 'very' even in the absence of the initial bridging context. In this way, the adverb becomes increasingly free to modify new adjectives without injecting its literal meaning as in (1-3b), effectively becoming bleached. Thus, the prediction we will test in order to evaluate our theory is as follows:

- **P1.** Rate of bleaching (for an adverb, over a given decade) is positively correlated with the similarity between an adverb and the adjectives modified by the adverb (henceforth SIMADJMOD).

4.2 Setup

We calculate rates of bleaching by taking the first derivative of extent of bleaching with respect to time, according to eq. 3:

$$\frac{d}{dt}(B(K,t)) = \frac{\Delta B}{\Delta t} = \frac{B(K,t+10) - B(K,t)}{10} \tag{3}$$

where $B(K,t)$ is rate of bleaching for an adverb K at time t according to one of the three bleaching metrics (SIMVERY, SIMLEX, BREADTH), giving us three different time series for rates of bleaching per adverb.

Since we are interested in examining how rate of bleaching over a given decade correlates with SIMADJMOD, the semantic relatedness between an adverb and the adjectives it modifies, we compute this variable (for a given adverb and decade) according to eq. 4:

$$\text{SIMADJMOD}(K,t) = \frac{\sum\limits_{a_i \in A_{K,t}} \mathtt{sim}(K,a_i)o(a_i)}{|A_{K,t}|} \tag{4}$$

where $A_{K,t}$ is the set of all adjectives modified by an adverb K at time t. Essentially, we take the average cosine similarity between an adverb and the adjectives it modifies, weighted by the odds of each adjective being modified (for a given decade).

4.3 Results

We present results using rates of bleaching computed from SVD embeddings (see Appendix C for results based on W2V embeddings). We find that our prediction is borne out: across all adverbs (both intensifiers and control), rate of bleaching over a given decade $D = [t_0, t_1)$ is positively correlated with SIMADJMOD at t_0 (the semantic relatedness between an adverb and adjectives modified at t_0), implying that at a given time, adverbs that modify semantically similar adjectives will bleach faster into intensifiers over the following decade. Lines of best fit from ordinary least squares regressions are shown in Fig. 3.

Moreover, what distinguishes intensifiers from non-bleaching control adverbs in our data is the variable SIMADJMOD: averaged across 1850-1990, SIMADJMOD is higher among the set of intensifiers compared to the set of control adverbs (Fig. 4). We further performed paired t-tests and found that SIMADJMOD is significantly higher for intensifiers than for the control adverbs (t = 7.3e+1, p<1e-20).[7]

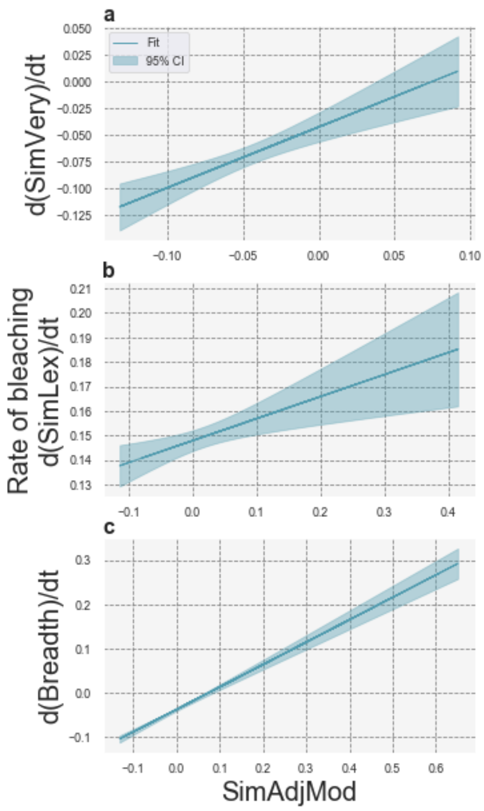

Figure 3. The more semantically similar an adverb is to the adjectives that it premodifies (the greater SIMADJMOD), the greater its rate of bleaching according to SIMVERY **(a)**, SIMLEX **(b)** and BREADTH **(c)**. Rates are computed using SVD embeddings and data are for all adverbs (intensifiers and control) at all years. Shaded areas show 95% confidence intervals.

[7]We also found the proportion of adjectives modified by an adverb relative to verbs to be significantly higher among intensifiers vs. control ($t = 4.4$, p<1e-04).

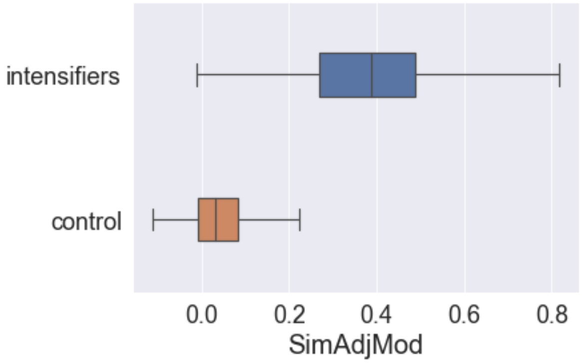

Figure 4. Intensifiers on average modify semantically more similar adjectives compared to control adverbs.

5 Discussion

In this work, we show how word embeddings and n-gram parse context can be used to model the semantic bleaching of manner adverbs into intensifiers. In particular, we empirically show that the bleaching of adverbs is associated with intuitive changes that have not previously been evaluated on large scale data: loss of root meaning, gain of target meaning, and increasing productivity. While our diachronic metrics may be biased by increasing corpus size over the years in our study, we find that the metrics SIMVERY, SIMLEX, and TYPEDIV still show significantly larger increases for the intensifiers compared to the control set. Thus, even though increasing corpus size presumably affects both wordsets equally, we have evidence to suggest that there are significant additional increases for intensifiers that may capture the fact that they are bleaching. We recommend that future researchers apply these metrics in conjunction with a control set (matched in frequency) when using other corpora subject to changes in size over time so that they may test for these significant relative differences between the bleaching and control words.

We also find that these two classes of adverbs can be distinguished in the absence of a control set when modeled using SIMLEX, an adverb's similarity to its root adjectival meaning. This metric also has the benefit over BREADTH of operationalizing a fundamental feature of bleaching that is not shared by other kinds of semantic change (e.g., metaphorical extension), as well as being generalizable (unlike SIMVERY) to cases of bleaching beyond manner adverbs becoming intensifiers. Thus, we recommend this metric to researchers interested in modeling bleaching more generally. We also show the utility of our methodology in evaluating explanatory hypotheses regarding how bleaching into intensifiers happens. We found that there is empirical evidence to support a reanalysis story: an adverb's tendency to modify adjectives that are semantically similar to itself is positively correlated with its subsequent rate of bleaching. This pathway of change is intuitive, as it is collocations such as *awfully disgusting* and *clearly obvious* that invite the re-interpretation of an adverb as a marker of emphasis, similar in function to an intensifier.

In future work, we are interested in refining BREADTH by normalizing for increasing corpus size as well as trying different weightings to capture the landscape of adjectives that an adverb modifies. It also remains an open question how generalizable our findings concerning bleaching of manner adverbs into intensifiers are. It would be interesting to see if other examples of adverb bleaching, such as the development of "moderators" (*slightly*, *hardly*, etc.) can be modeled as reanalysis. Another under-explored example of adverb bleaching concerns the development of maximizing adverbs into reinforcing adverbs. Beltrama and Staum Casasanto (2017) study the change undergone by *totally*, but the larger tendency remains unexplored.

Furthermore, among English adverbs, there are many other semantic factors that have potential effects on bleaching. For example, Sweetser (1989) suggests that words which explicitly highlight semantic facets of the source domain that cannot be mapped onto the target domain are unlikely candidates for grammaticalization as they require "active suppression" of the foregrounded meanings.[8] It would be interesting to study the bleaching of intensifiers with this question in mind—for example, the adverb *vanishingly* occurs in contexts like *vanishingly small* and *vanishingly rare* which are well-suited for reanalysis, but for *vanishingly* to be understood as a generic intensifier would also require suppression of its meaning of smallness.

Acknowledgments

We thank Daniel Lassiter, Vivek Kulkarni, and anonymous reviewers for their helpful feedback.

[8] For example, the verb *lumber* is an unlikely candidate for undergoing the change from motion verb to tense marker because it explicitly encodes rate and manner of motion, compared to a verb like *go*.

References

Andrea Beltrama and M Ryan Bochnak. 2015. Intensification without degrees cross-linguistically. *Natural Language & Linguistic Theory*, 33(3):843–879.

Andrea Beltrama and Laura Staum Casasanto. 2017. Totally tall sounds totally younger: Intensification at the socio-semantics interface. *Journal of Sociolinguistics*, 21(2):154–182.

Dwight Bolinger. 1972. *Degree Words*. Paris: Mouton.

Michel Bréal. 1964. *Semantics: Studies in the science of meaning*. New York: Dover.

Joan Bybee. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5):425–455.

Joan L Bybee, Revere Dale Perkins, and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*, volume 196. The University of Chicago Press: Chicago.

Mark Davies. 2010-. *The Corpus of Historical American English (COHA): 400 million words, 1810-2009*. Available online at https://corpus.byu.edu/coha/.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145, Copenhagen, Denmark. Association for Computational Linguistics.

Nicholas Evans and David Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76:546–592.

Lea Frermann and Mirella Lapata. 2016. A Bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics*, 4:31–45.

Yoav Goldberg and Jon Orwant. 2013. A dataset of syntactic-ngrams over time from a very large corpus of English books. In **SEM*.

Kristina Gulordava and Marco Baroni. 2011. A distributional similarity approach to the detection of semantic change in the Google books ngram corpus. In *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 67–71, Edinburgh, UK. Association for Computational Linguistics.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016a. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2116–2121, Austin, Texas. Association for Computational Linguistics.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016b. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Paul J Hopper and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web*, pages 625–635. International World Wide Web Conferences Steering Committee.

Yuri Lin, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman, and Slav Petrov. 2012. Syntactic annotations for the google books ngram corpus. In *Proceedings of the ACL 2012 System Demonstrations*, pages 169–174, Jeju Island, Korea. Association for Computational Linguistics.

Gunter Lorenz. 2002. Really worthwhile or not really significant? A corpus-based approach to the delexicalization and grammaticalization of intensifiers in Modern English. *New reflections on grammaticalization*, 49:143.

Yiwei Luo and Yang Xu. 2018. Stability in the temporal dynamics of word meanings. In *CogSci*.

Ronald Macaulay. 2006. Pure grammaticalization: The development of a teenage intensifier. *Language Variation and Change*, 18(3):267–283.

George Miller. 1998. *WordNet: An Electronic Lexical Database*. MIT press.

Sunny Mitra, Ritwik Mitra, Martin Riedl, Chris Biemann, Animesh Mukherjee, and Pawan Goyal. 2014. That's sick dude!: Automatic identification of word sense change across different timescales. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1020–1029, Baltimore, Maryland. Association for Computational Linguistics.

Marcin Morzycki. 2008. Adverbial modification of adjectives: Evaluatives and a little beyond. *Event structures in linguistic form and interpretation*, 5:103.

Rick Nouwen. 2011. Degree modifiers and monotonicity. In *Vagueness and language use*, pages 146–164. Springer, Palgrave.

Mark Pagel, Quentin D Atkinson, and Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449(7163):717.

Carita Paradis. 1997. *Degree modifiers of adjectives in spoken British English*, volume 92. Lund: Lund University Press.

Hans Peters. 1994. *Degree adverbs in Early Modern English*, volume 13. Berlin: Walter de Gruyter.

Christian Ramiro, Mahesh Srinivasan, Barbara C. Malt, and Yang Xu. 2018. Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences*, 115(10):2323–2328.

Alex Rosenfeld and Katrin Erk. 2018. Deep neural models of semantic shift. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 474–484, New Orleans, Louisiana. Association for Computational Linguistics.

Skipper Seabold and Josef Perktold. 2010. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

John Simpson, Edmund SC Weiner, et al. Oxford English Dictionary online. *Accessed October 2018*.

Gustaf Stern. 1931. *Meaning and change of meaning*. Bloomington: Indiana University Press.

Eve Sweetser. 1989. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*, volume 54. Cambridge: Cambridge University Press.

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

Yang Xu and Charles Kemp. 2015. A computational evaluation of two laws of semantic change. In *CogSci*.

A Full set of 250 intensifiers

abnormally	abominably	absolutely
abundantly	abysmally	actually
acutely	adamantly	aggressively
alarmingly	amazingly	amply
annoyingly	astonishingly	astronomically
atrociously	awfully	basically
beautifully	bitterly	blatantly
breathtakingly	brutally	categorically
clearly	cloyingly	colossally
comically	completely	considerably
conspicuously	copiously	crazily
criminally	curiously	dangerously
decadently	decently	decidedly
deeply	defiantly	definitely
delectably	deliciously	delightfully
depressingly	desperately	devastatingly
disastrously	disconcertingly	disgustingly
dismayingly	distinctly	distressingly
disturbingly	dizzyingly	doubly
dramatically	dreadfully	egregiously
embarrassingly	empatically	endlessly
enormously	enthusiastically	entirely
epically	especially	evidently
exceedingly	excellently	exceptionally
excessively	excruciatingly	exorbitantly
extensively	extraordinarily	extravagantly
extremely	exuberantly	fairly
fiercely	firmly	fortunately
frightfully	frustratingly	fully
fundamentally	furiously	genuinely
gorgeously	greatly	grievously
grossly	handsomely	harshly
heavily	hellishly	hilariously
honestly	horribly	horrifically
hugely	hysterically	immensely
immoderately	impossibly	impressively
improperly	inappropriately	inconveniently
indecently	indescribably	inestimably
inexcusably	inexplicably	infinitely
insanely	intensely	intimately
intolerably	justly	laughably
lavishly	legitimately	liberally
literally	ludicrously	luxuriously
madly	magically	magnificently
majorly	marginally	markedly

marvellously	massively	mightily
mind-blowingly	mindlessly	miraculously
miserably	monstrously	mysteriously
needlessly	nicely	notably
noticeably	notoriously	objectively
obnoxiously	obscenely	offensively
outrageously	outstandingly	overbearingly
overpoweringly	overtly	overwhelmingly
painfully	particularly	passionately
pathetically	perfectly	phenomenally
pleasantly	profusely	prominently
purely	radically	reasonably
recklessly	regretfully	regrettably
relentlessly	reliably	remarkably
revoltingly	richly	savagely
scarily	seriously	severely
shamelessly	sharply	shockingly
sickeningly	significantly	simply
sincerely	sinfully	solidly
sorely	spectacularly	splendidly
startlingly	strangely	strikingly
strongly	stunningly	stupendously
stupidly	substantially	superbly
supremely	surpassingly	surprisingly
terribly	terrifically	thankfully
thoroughly	threateningly	totally
tragically	tremendously	truly
unapologetically	unbearably	uncomfortably
uncommonly	uncontrollably	undeniably
undoubtedly	unequivocally	unexpectedly
unfortunately	unjustly	unmistakably
unnecessarily	unnervingly	unpleasantly
unquestionably	unreasonably	unsettlingly
unspeakably	unusually	unutterably
utterly	vastly	veritably
vigorously	violently	virtually
visibly	weirdly	wholeheartedly
wholly	wickedly	wildly
woefully	wonderfully	worryingly

B Full set of 178 control adverbs

abruptly	accordingly	accurately
actively	adequately	allegedly
alternatively	angrily	annually
apparently	appropriately	approximately
automatically	badly	barely
bitterly	briefly	broadly
carefully	comfortably	commonly
comparatively	consequently	consistently
constantly	continually	continuously
conversely	correctly	currently
daily	deliberately	differently
directly	duly	easily
easily	economically	effectively
efficiently	equally	essentially
eventually	exactly	exclusively
explicitly	finally	financially
firstly	formally	formerly
frankly	freely	frequently
generally	gently	gradually
happily	hastily	historically
hopefully	ideally	immediately
importantly	incidentally	increasingly
independently	indirectly	individually
inevitably	initially	instantly
invariably	ironically	jointly
kindly	lately	legally
lightly	locally	loudly
mainly	mentally	mostly
namely	neatly	necessarily
newly	normally	obviously
occasionally	officially	openly
originally	partially	partly
permanently	personally	physically
politically	poorly	positively
possibly	potentially	practically
precisely	predominantly	presently
presumably	previously	primarily
principally	privately	probably
promptly	properly	publicly
quickly	rapidly	rarely
readily	recently	regularly
reportedly	repsectively	rightly
roughly	sadly	safely
secondly	seemingly	separately
sexually	shortly	silently

similarly	simultaneously	slowly
smoothly	socially	softly
solely	specifically	steadily
strictly	subsequently	successfully
suddenly	sufficiently	supposedly
swiftly	technically	temporarily
tightly	traditionally	typically
ultimately	urgently	usually
vaguely	weakly	widely

C Diachronic correlations for W2V embeddings-based rates of bleaching

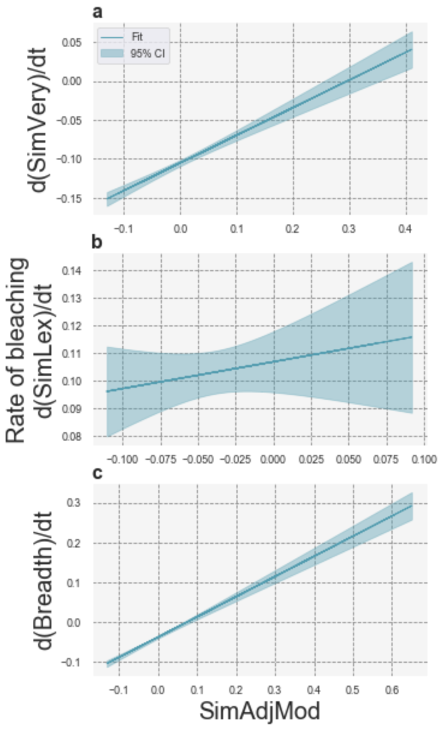

Figure 5. The more semantically similar an adverb is to the adjectives that it premodifies, the greater its rate of bleaching by all three metrics. Rates are computed using HistWords W2V embeddings and data are for all adverbs (intensifiers and control) at all years.