Sentiment in English Degree Adverbs:
a Perspective from Collostruction and Network Analysis

HSIEH, WALTER

Taipei Fuhsing Private School 11th-grade

Instructor: Su Hung Kuan

September 2023

1. Introduction

Language is a realm of rich possibilities, offering speakers a spectrum of linguistic choices to navigate diverse social situations. These linguistic alternatives, while similar, develop unique usages in everyday discourse, allowing speakers to accurately adjust their expressions. Understanding the principles underlying these choices can be helpful for the education of English adverbs and can even shed light on how similar concepts are represented in human minds.

This study delves into a specific facet: degree adverb constructions (DAC) such as VERY, QUITE, TOTALLY, etc. Degree adverb constructions are pivotal in conveying different subjectivity, enabling speakers to include varied intensities of emotions in their sentences. Through quantitative methods (i.e. correlation analysis, sentiment analysis, and network analysis), this study aims to uncover their intricate nuances and sentiment implications and argues that in spite of their near-synonymous meanings, DACs differ in terms of their attracted adjective, which reveals different semantic polarities carried by these constructions.

2. Literature Review

2.1 Degree adverbs in spoken Mandarin

Huang & Chen (2022) explored four degree adverbs, *tài*, *mán*, and *chāo* in Mandarin, using the corpus in TalkBank. The corpus-based behavioral profile method discerns two super-clusters: *chāo-mán* and *tài-hěn*. *Chāo* and *mán* differ in pragmatic sentiments, nominalization co-occurrences, and lexicalization productivity. Similarly, *tài* and *hěn* vary in *-le* particle co-occurrences, predicate functions, and associated head noun semantics. Bridging empirical analysis and established linguistic inquiry, this study contributes to comprehending the more subtle usage of degree adverb constructions in spoken Mandarin.

2.2 Classifying the Natural Sentiment Strength of Degree Adverbs

Darwich et al. (2019) focus on a new approach that captures sentiment direction and strength. Traditional models often categorize text into discrete classes, but this study's method places sentiment on a spectrum between positive and negative, providing a more accurate sentiment representation. This study introduces the 'Natural' Sentiment Strength (NSS) model. Unlike statistical methods, NSS relies on human-assigned meanings found in glosses, semantic networks, and predefined adverbs to quantify sentiment strength naturally. It can be seen through this study that degree adverbs actually vary quite a lot, and using sentiment to analyze these adverbs would be a great way to classify them.

2.3   Collostructional Analysis and Network Analysis

This study introduces the method of collostructional analysis, a methodology developed by Gries and Stefanowitsch in 2003 which is used to measure the attraction, or correlation, of words towards each other. This corpus method made it possible to quantify the attraction and repulsion between the DACs and their subsequent adjectives, in the case of this study. This method also proves to be better than using raw frequency to predict accuracy, as it takes into account the total number of a word versus the number of the word that is attached to its counterpart.

Apart from using correlation numbers from collostructional analysis, another essential tool used in this study is network analysis. Network analysis allows for the visualization of the data collected via collostructional analysis. It allows researchers to explore how a group of near-synonymous words can be sub-categorized by visualizing the semantic distance among words. It serves as a good tool to validate the grouping based on the sentiment of the co-occurring adjectives.

2.4   Research Questions

To sum up, the study in 2.2 demonstrates the importance of sentiment analysis on adverbs. However, few studies have explored the sentiment of DAC. Most sentiment dictionaries such as QDAP, Harvard IV classified DACs as sentimentally neutral without any semantic prosody. However, in the sentence "The food is *quite/very/totally/completely* good", the usage of different DACs reveals different levels of acceptance among English speakers, which implies that these DACs could have different attraction toward different types of adjectives. Huang & Chen (2022) point out that even though these DACs seem near synonymous, it doesn't mean that they are replaceable. Apart from differences in usage, polarity might also affect the usage of DACs in their sentences. The concept of semantic polarity behind DAC is what the present study wants to investigate further. In other words, this study aims to use a corpus-based approach to analyze the sentiments of different degree adverbs and use the sentiment of subsequent adjectives as sources to categorize DACs.

3. Methodology

3.1   Data Extraction and Filtering

The British National Corpus, the corpus used in this study, is a corpus developed by linguists from Oxford University and Lancaster University. Among the 100 million words in the corpus and the millions of sentences in the corpus, about 10% is spoken English of the

late 20th century to the 21th century, collected in conversations recorded in everyday life or on television. For this study, the spoken corpus is used because the purpose of this study is to determine how people use these DACs in everyday social situations instead of when writing on pen and paper. After choosing the BNC spoken corpus, this study uses the NTNU corpus resources website for advanced filtering and obtaining of the actual sentences for each of the DACs. By entering specific parameters into the online corpus, it was possible to filter out the top 10 frequent DACs as well as their subsequent adjectives and the concordance lines.

As for the analytic tools, the present study used collostructional analysis to extract and filter the subsequent adjectives for analysis. By looking at their sum of absolute deviation (sumabsdev), the adjectives that had lower sumabsdev values were the least deviant from expected, meaning that their behavior was closer to random chance, and hence were deleted from the observed data. Afterward, utilizing a self-developed R script, sentiment analysis was used to determine the sentiment leanings of each word, which helped group the DACs into the groups presented in section 4. Finally, network analysis is performed to validate the grouping observed from sentiment analysis, and to observe more possible ways of categorization.

4. Results and Analysis

In this section, we will be grouping the DACs into their own respective groups based on Table 4.1:

(1) **Polarizing DAC with specific preference (Negative DACs and Positive DACs):** This group includes DACs with neutral sentiment below 50% and either positive or negative sentiment is lower than 20% (refer to Table 4.1). Typically, this type of construction would be more likely to attract either more positive adjectives in positive DACs or more negative adjectives in negative DACs.

(2) **Polarizing DAC without specific preference:** This group includes DACs with neutral sentiment below 50% and neither positive nor negative sentiment is lower than 20% (refer to Table 4.1). Typically, this type of construction would be more likely to attract adjectives with sentiment but does have an obvious tendency toward either positive or negative adjectives.

(3) **Neutral DAC:** This group includes DACs with neutral sentiment over 50% (refer to Table 4.1). Typically, this type of construction would be more likely to attract adjectives without sentiment.

Table 4.1: The sentiment distribution of subsequent adjectives for each degree adverb (in percentages)

	Type	Negative	Neutral	Positive
completely	Polarizing DAC with specific preference	45.93	48.79	5.27
totally		61.54	32.48	5.98
quite		7.59	27.23	65.18
very		18.52	44.44	37.04
particularly	Polarizing DAC without specific preference	39.57	25.18	35.25
absolutely		31.76	22.44	45.80
really		45.00	14.09	40.91
so	Neutral DAC	36.56	52.42	11.01
too		29.96	64.56	5.49
all		20.05	54.20	25.75

*All numbers are rounded to the nearest hundredth (%)

4.1 Polarizing DAC with specific preference

4.1.1 Negative DAC– *completely*, *totally*

4.1.1.1 Completely

different	anonymous	irrelevant	new	confidential	wrong	separate	inadmissible	uncalled	empty
43.1158	26.7066	23.0235	17.9683	17.9039	16.6836	15.0637	14.1688	14.1688	14.0290

The table shows top 10 adjectives that have the strongest attraction to *completely*. Based on the table above, there are no positive adjectives in the top 10 subsequent adjectives relative to *completely*, with the most prominent by far being *different*, which indicates that this category repels positive adjectives. Therefore, it might explain why it would be more acceptable to say something is *completely wrong* rather than *completely right*. Another point that should be noted is that although it is categorized as a negative DAC. From Table 4.1, we can see that compared with DAC *totally*, the tendency toward negativity is less strong, which therefore can only for the present study to claim that this construction tends to "repel" positive adjectives rather than it tends to attract negative adjectives. Nevertheless, since a large proportion(45%) of co-occurring adjectives are negative. The construction would in some cases convey the subjectivity that the speaker holds a negative evaluation toward the event as in (1).

(1) First of all leaving such a beautiful country and leaving all my family coming over
    here to a foreign environment to a different world. The football world is a *completely*
    different world even it's different to what people would be used to you know here.

In (1), we can see how this *completely* modifies the neutral adjective *different*, and the
speaker tried to use *completely different* to highlight the difficulty of getting used to a new
environment, which in his description is in contrast to his *beautiful country*.

4.1.1.2 Totally

different	unnecessary	wrong	unacceptable	incompetent	illiterate	dependent	unexpected	passive	inadequate
42.0869	23.5233	20.2536	19.5531	13.9343	13.1874	12.6960	12.4158	12.4158	11.2908

Totally is like *completely*, in that it barely has any positive adjectives that follow after it.
Another point worthy of noting is that compared with completely, this construction has
attracted more negative adjectives in top 10 as in the above table; in addition, from Table 4.1,
the proportion of negativity takes up to 60% . We might be able to infer that This construction
might have a stronger negative subjectivity than completely. Besides, the adjectives attracted
to this construction are affixed with a negation prefix, which further demonstrates that this
construction might want to emphasize the fact that the head nouns modified by these
adjectives are not capable of fulfilling the quality as expected by the speaker. For example, in
(2), the speaker tries to use *totally unnecessary* to show that things are going opposite against
his expectations as we can see from the last line.

(2) This is frequently said. We've added on. In our Information Technology Department
    we've now got thirty two posts. This is *totally unnecessary*. We talk about equal
    opportunities, we all believe in equal opportunities. It is right and proper that
    everyone should have the same opportunity, but this can be controlled by one officer
    in my belief.

(3) When your reading I closed my eyes and I got a different feeling. Yes from your and I
    can visualise, something *totally different*. But is that what it should be? Yes, that's
    what, yes that's what I hoped for because I, I feel everybody's interpretation yes is
    different,

Even in (3), *totally* is taking on a neutral adjective *different*. It still retains the interpretation
we gain from (2) going against one's expectation or belief, and in this sentence, the speaker
uses this construction to show that the speaker thinks that the situation is going against the
listener's belief.

4.1.2 Positive DAC – *very, quite*

4.1.2.1 Very

good	difficult	important	little	few	similar	careful	large	small	grateful
14.9611	12.5851	11.5943	8.8358	8.7642	8.2007	7.6963	6.5813	5.5850	5.5814

Very has a lot of possible subsequent adjectives because of how common it is. Therefore, the correlation scores of the top 10 are all really low compared to the correlation scores of adjectives for other adverbs. However, it is still slightly leaning towards a positive and neutral sentiment, which is why it is grouped in this category. Although its top 10 correlations mostly consist of neutral adjectives, it can make these neutral adjectives presented in a positive way.

(4) And as we wait and watch the findings are *very much* those of Paul in Bristol, that a great deal of these patients get better, and only *very few* fall into problems.

(4) is the perfect example of the phenomenon mentioned above, as *very* helps the neutral adjectives to be presented in a more positive context in both *very much* and *very few*.

4.1.2.2 Quite

sure	happy	clear	right	honest	interesting	possible	likely	normal	willing
25.9303	24.7653	16.5116	16.1133	15.4505	10.7263	10.2592	9.7335	6.3168	6.3016

In the top 10 collocation adjectives for *quite*, it is evident that most adjectives here have a positive sentiment, with the neutral adjectives being able to pass as positive after you really look at them as well.

(5) On the basis that the recommendations were acceptable would you be happy to proceed next time? Yes I think that's *quite possible*.

In this example, *quite possible* suggests that there is a more likely chance that the recommendations would be acceptable, giving the phrase a more positive sentiment overall.

4.2 Polarizing intensifiers – *particularly*, *really*, and *absolutely*

4.2.1 Particularly

interested	hazardous	sensitive	vulnerable	addictive	backward	unique	conscious	keen	important
25.9303	24.7653	16.5116	16.1133	15.4505	10.7263	10.2592	9.7335	6.3168	4.8606

For *particularly*, there is no specific leaning toward either negative or positive adjectives. However, there is a considerably less quantity of neutral adjectives in the top 10, suggesting that *particularly* is a more polarizing adverb that focuses on emphasizing the sentiment of

adjectives with already existing sentiment, instead of giving sentiment to comparatively neutral adjectives.

(6) [A]ll the Rule 43 prisoners, who were the ones who were *particularly vulnerable*, are now accounted for, and I mean by that they have all come out of the jail in the last few days and unfortunately one of them, as you know, has died from his injuries.

In this sentence, it is clear that *particularly* is really making the negative sentiment of *vulnerable* more clear, as the one of the prisoners' vulnerability eventually led to their deaths.

4.2.2 Really

nice	weird	lovely	worth	funny	good	horrible	nasty	cool	disgusting
14.8751	13.2451	13.0008	11.5065	11.2481	10.4069	9.8681	7.3925	7.2270	5.2718

In the top 10 of *really*'s adjective collocations, there are no neutral adjectives, making it an even better example for this category. Like *very*, it is also very common to use the word *really*, hence the lower correlation scores to all adjectives. The sentiment distribution within the top 10 is quite even, suggesting that *really* doesn't carry any sentiment, and just strengthens the sentiments of other adjectives.

(7) And off they go. It's *really funny*. I'd love to live there.

One of the more curious cases is *funny*, which has a negative sentiment in the dictionary. In this sentence, however, it is showing really positive sentiment without the help of *really*.

4.2.3 Absolutely

essential	right	marvelous	brilliant	perfect	certain	wonderful	fantastic	gorgeous	vital
26.2276	24.9383	23.2819	22.1065	21.6326	21.3146	20.3002	17.6370	16.8245	15.4734

At first glance, *absolutely* should have been sorted into the category of positive degree adverbs. However, according to Table 4.1, the amount of positive sentiment adjectives' scores combined is nearly the same as the negative score, suggesting that maybe after the top 10, most adjectives correlated are negative. Instead of having adjectives that are gradable, like *really* and *particularly*, however, its list consists of more "black and white" adjectives, like *certain*, *right*, and *perfect*.

(8) I think it's *absolutely essential* that the staff advised this.

Evidently, *absolutely* does not really alter the sentiment of this neutral adjective, which is really what separates this group's adverbs from the others, where adverbs would slightly alter the sentiments of neutral adjectives.

4.3 Neutral DAC– *too*, *so*, and *all*

4.3.1 Too

bad	late	much	many	big	hot	early	long	juicy	young
29.9801	29.6473	28.3920	18.5797	13.2622	10.9634	10.8517	9.6061	9.1466	7.3352

Above are the top 10 subsequent adjectives with respect to the degree adverb too in the context of correlation of collocations. Evidently, except for the one positive adjective and the two negative adjectives that are further down the list, the top 10 is mostly dominated by neutral adjectives. Some of the similarities that can be seen in these top 10 adjectives related to too are the fact that they mostly consist of subjective adjectives, including some that are sensory details like *juicy*, *big*, and *hot*.

(9) So the people were most concerned that women's morale should be kept up and that one shouldn't worry *too much*.

In (9), *much* is a neutral word, but *worry* is rather negative, making this *too* in a place of negative sentiment

(10)   Alas, *too well* you are acquainted, sir, with the distressed condition of my heart.

In (10), *too* is used in a more neutral context.

4.3.2 So

many	much	funny	stupid	bloody	long	good	bad	cute	sorry
40.8888	18.6691	12.6806	10.2087	7.6842	6.6682	6.0230	5.7646	5.7116	5.6991

So is like *too,* but with more negative words in the top 10. This doesn't affect the general sentiment of the adjectives subsequent to *so. Many* and *much* combined alone have a score of 60, while the rest of the negative adjectives only have scores of 12 or lower. Like *too*, the top 10 list includes more subjective adjectives. It is different in that the adjectives in this list aren't as based on sensory details as the top 10 adjectives for too.

(11)   And I just liked reading it *so much* that I thought you'd like to read it too.

In (11), *much* is a neutral word, but the situation makes this sentence more positive.

4.3.3 All

ready	white	yellow	identical	sticky	dead	wrong	round	asleep	familiar
48.16434	18.4057	15.1568	13.2376	13.0148	12.4685	11.5031	11.4089	11.3998	10.9268

All, while it may seem less common, is actually similar to *so* and *too*. It, however, precedes very different types of adjectives, with more adjectives describing the state of something rather than a subjective description, like *all ready* or *all asleep*.

(12)   Are you *all ready* for Christmas, or no?

In (12), *all* doesn't really provide much sentiment, making the sentence neutral as it is.
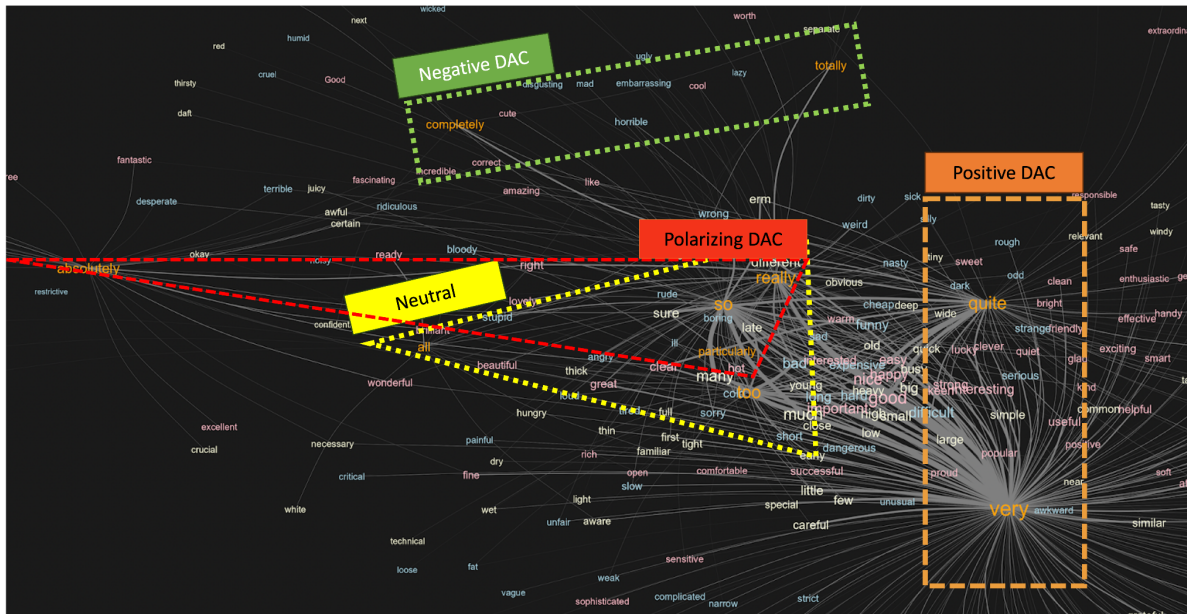
5. Network Analysis

Figure 5.1 Network analysis graph of DACs and subsequent adjectives

After using a self-written R script for network analysis to group these DACs by their correlation to one another and their most related adjectives, we are able to validate the groups that we formed based on the sentiment of the attracted adjectives in the previous section. In network analysis, we can see DACs marked in orange scatter across the figures. In addition, there are also other words in pink/yellow/blue. They are the words that cooccur with these DACs and if an adjective cooccurs with the DAC once, there would be a line, so-called edge, connecting the two words. The stronger the attraction based on collostructional analysis between a DAC and an adjective, the thicker the edge would be. Therefore, network analysis is powerful in terms of depicting the complex ecosystem of words and constructions.

From Figure 5.1, it is clear that the positive and negative DACs were quite separated from the others, both being in their own framework on the side. However, the neutral and polarizing DACs are a lot more overlapping with one another, implying that the adjectives that they use, although seemingly different based on the sentiment analysis in Table 4.1, are actually quite similar. Moreover, within these two groups, *absolutely* and *all* are the specifically different DACs that veer far away from the other two DACs. This raises an intriguing potential future

research question: Why is it that these two DACs, despite their similar sentiment, are so different from the other DACs of their group?

6. Conclusion

This research paper reviews the general usage of DACs in spoken English, using data from the British National Corpus. By correlation numbers obtained from collostructional analysis, this study was able to group the top 10 used DACs in spoken English into four general categories based on the sentiments of their most correlated subsequent adjectives, more specifically into the group of neutral, positive, negative, and polarizing intensifiers. Future studies could add on to this study by including a study that would group adjectives into more groups instead of just three groups based on their sentiment. For example, one phenomenon discovered in this study was that certain adverbs had the capability of turning some originally not so gradable adjectives into gradable adjectives. Gradability could be one of the many properties to also include in grouping these adverbs in future studies. Of course, in the future, one could also use similar techniques to replicate, but instead, using written English. This could maybe differentiate some of the more formal adverbs from the more casual adverbs that are used in everyday conversations.

References

Darwich, M., Mohd Noah, S. A., Omar, N., Osman, N., & Said Ahmad, I. (2020, July 24). Quantifying the Natural Sentiment Strength of Polar Term Senses Using Semantic Gloss Information and Degree Adverbs. Journal of Advances in Information Technology, Vol. 11, No. 3, 109-118. https://doi.org/10.12720/jait.11.3.109-118

Huang, P.-W., & Chen, A. (2022, November 29). Degree adverbs in spoken Mandarin: A behavioral profile corpus-based approach to language alternatives. Concentric. Studies in Linguistics, 48, 285-322. https://doi.org/10.1075/consl.22002.che

Stefanowitsch, A., & Gries, S. (2003, December 31). Collostructions: Investigating the interaction of words and constructions. International Journal of Corpus Linguistics, 8(2). https://doi.org/10.1075/ijcl.8.2.03ste