

ATINER CONFERENCE PAPER SERIES No: LNG2013-0840

**Athens Institute for Education and Research**

**ATINER**



**ATINER's Conference Paper Series**

**LNG2013-0840**

**Study on the Sentiment Polarity Types  
of Collocations for *too* and *very***

**Mohamed Yassine Frej**

**DICORA Researcher**

**DICORA/ Hankuk University of Foreign Studies,  
Korea**

**Jeesun Nam**

**Full Professor**

**DICORA/ Hankuk University of Foreign Studies  
Korea**

ATINER CONFERENCE PAPER SERIES No: LNG2013-0840

Athens Institute for Education and Research  
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece  
Tel: + 30 210 3634210 Fax: + 30 210 3634209  
Email: [info@atiner.gr](mailto:info@atiner.gr) URL: [www.atiner.gr](http://www.atiner.gr)  
URL Conference Papers Series: [www.atiner.gr/papers.htm](http://www.atiner.gr/papers.htm)

Printed in Athens, Greece by the Athens Institute for Education and Research.  
All rights reserved. Reproduction is allowed for non-commercial purposes if the  
source is fully acknowledged.

**ISSN 2241-2891**

23/1/2014

## An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

Dr. Gregory T. Papanikos  
President  
Athens Institute for Education and Research

This paper should be cited as follows:

**Frej, M.Y. and Nam, J. (2013) "Study on the Sentiment Polarity Types of Collocations for too and very" Athens: ATINER'S Conference Paper Series, No: LNG2013-0840.**

## **Study on the Sentiment Polarity Types of Collocations for *too* and *very***

**Mohamed Yassine Frej**  
**DICORA Researcher**  
**DICORA/ Hankuk University of Foreign Studies,**  
**Korea**

**Jeesun Nam**  
**Full Professor**  
**DICORA/ Hankuk University of Foreign Studies**  
**Korea**

### **Abstract**

This study is a corpus-based approach conducted to compare the two amplifiers *too* and *very* in terms of what sentiment polarity types can be attributed to the words in collocation with these adverbs. These adverbs are generally known as intensifying a given meaning of modified words, notably adjectives, verbs or nouns. However, based on the empirical examination of the British National Corpus (BNC), we assume that the prominent sentiment polarity type of the words collocated with *very* is dissimilar to that of the words collocated with *too*. For instance, while the adjectives collocated with *very* expresses rather a positive sentiment polarity like in ‘*The soup is very **tasty***’, those appeared with *too* mostly expresses a negative sentiment polarity like in ‘*This car is **too expensive***’. In most previous works, it is asserted that it is due to the particular semantic function of the adverb *too*, since no matter what semantic types of adjectives are, the sequences such as *too small*, *too bright* or *too young* express a negative evaluation. Nevertheless, we advocate that this interpretation does not result from the special function of *too*, but from the semantic orientation itself of the words collocated with these adverbs. The experiment we performed in this study reveals that the adjectives of negative polarity co-occur with *too* 7 times more frequently than those of positive polarity. On the contrary, the adjectives of positive polarity co-occur with *very* 2.6 times more frequently than those of negative polarity. The distribution of part-of-speech in the position collocated with these adverbs, the proportion of words with positive/negative/neutral polarity and the corpus types such as formal/written texts or informal/spoken texts are discussed as well to clarify the empirical differences of these adverbs.

**Keywords:**

**Corresponding Author:**

## Introduction

This study is a corpus-based approach conducted to compare the two amplifiers *too* and *very* in terms of what sentiment polarity types can be attributed to the words in collocation with these adverbs. In human languages, more can be learned about the differences between certain words through the words they collocate with. The latter are referred to, in the literature, as collocations. Firth (1957) emphasized the importance of both linguistic and situational contexts for the description of human languages. “You shall know a word by the company it keeps.” (p. 195).

The abundance of large collections of spoken and written corpora in electronic form and the possibility to investigate those texts made by the development of technologies in corpus linguistics over the last years laid the path towards better understanding of how words are arranged in languages. Thus, the nature of vocabulary and its contribution to the existing semantics and syntax-based theories are reconsidered now.

Adverbs of degree and their collocations have been extensively examined in the literature. One of the first studies is Quirk et al (1985). The framework used was essentially a semantic one, by which amplifiers also called intensifiers, such as *absolutely*, *completely*, *really* and *very*, were considered to express degrees of increasing intensity, whereas downtoners, such as *rather*, *a bit*, *somewhat* and *quite*, were considered as scaling the sense of an adjective downward from an assumed norm, often used as a softening device in cases where the adjective sounds too strong when left unmodified.

Based on a large (40-million words) corpus of American and British spoken and written English is another important corpus-based study of amplifier-adjective by Biber, Johansson, Leech, Conrad, and Finegan’s (1999). The description shows that the most frequent amplifiers immediately preceding adjectives in British English conversation are *very*, *so*, *really*, and *too* followed by *absolutely*, *bloody*, *damn*, *real*, *completely*, and *totally*. In American English conversational genres the distribution is similar, except for *bloody*, which is infrequent, and *real*, which is much more frequent than in British English. *Extremely*, *highly*, *entirely*, *fully*, *incredibly*, *perfectly*, *strongly*, and *terribly* also occur frequently in both regional varieties, especially in written, academic genres. Biber et al. (1999, p. 545) shows that the most frequent amplifier-adjective collocations in British English conversation are *very good*, *very nice*, *really good*, *really nice*, and *too bad*, whereas in American English conversation *really good*, *too bad*, *very good*, *real good*, *real quick*, *really bad*, *really nice*, *too big*, and *very nice* are the most frequent. Also, they suggested that they somewhat tend to be used interchangeably in certain contexts despite their different meanings. The suggestions of this study are interesting and need to be speculated using a different corpus.

Due to the growing interest in sentiment analysis of online reviews people write for all types of products and services, a huge amount of work has been devoted to the analysis of subjective expressions. Recent research focuses on assigning a polarity such as positive or negative to subjective expressions (i.e.

adjectives, adverbs, verbs...) which contain opinions, emotions or sentiments. Although those subjective expressions are indispensable to determine the sentiment polarity of utterances, the overall sentiment polarity of sentences might be affected by another type of devices referred to as polarity shifting markers. Negation markers such as *not* are one type of those polarity shifting markers (Kennedy & Inkpen 2006). For example, in the sentence *the car is not comfortable*, the polarity of the word *comfortable* is positive, whereas the polarity of the whole sentence is reversed because of the negation marker *not*. In fact, valence shifters (Polanyi & Zaenen, 2004), the terms that can change the semantic orientation of another term such as negation, should be taken into consideration. Among diverse valence shifters the authors suggested, negatives and intensifiers seem to be the most obvious shifters, and thus Kennedy & Inkpen (2006) took these particular devices into account in order to improve the accuracy of their sentiment classification system. However, we advocate in this paper that, unlike negation markers, only few of the occurrences of intensifiers like *too* and *very* are used as valence shifters, and that even when those intensifiers exist in sentences conveying a negative evaluation, this evaluation does not result from the special function of *too* or *very*, but from the semantic orientation itself of the words collocated with these adverbs. Moreover, not much attention has been given to uncover how those intensifiers are different from each other, and what contexts are required so that the interpretation of the adverb is determined. Therefore, the distribution of part-of-speech in the position collocated with these adverbs, the proportion of words with positive/negative/neutral polarity and the corpus types such as formal/written texts or informal/spoken texts are discussed as well to clarify the empirical differences of these adverbs.

## Method

The BNC (Leech, Rayson & Wilson, 2001) is a 100-million-word structured collection of spoken and written texts. The corpus was compiled by a consortium of universities, publishers, and the British government in the 1990s to be representative of the spoken and written English used in Britain at the end of the 20<sup>th</sup> century. The BNC includes 90 million words of written English from eight genres and 10 million words of spoken English from four social class groupings, collected in 38 locations in the United Kingdom. Therefore, the BNC was chosen because of its considerable size and also because it incorporates a robust grammatical tagger which facilitates the retrieval and analysis of words.

Adjectives associated with the amplifiers (*too* and *very*) were the focus of the present study. *Very* and *too* were selected largely because they are among the most frequent amplifiers in the corpus respectively (see table 1).

**Table 1.** *Frequency of too/very in the BNC*

	<i>Very</i>	<i>Too</i>
Frequency	119,572	67,004

The questions are these highly frequent adverbs similar or different from each other in contributing to sentiment polarity? How often do they express a negative opinion? What kind of collocations are they associated with? To answer these questions, we shall examine the part of speech (POS) types that occurred in the right context of *very* and *too* in the BNC corpus as well as the evaluation types of these intensifiers in this corpus.

### Analysis

#### *POS collocation Types*

What we first notice from the occurrences in the table above is that the adverbs under study tend to be associated with a given POS more than another. This is shown clearly in the table below:

**Table 3.** *POS Distribution of the collocations of too/very*

	<i>too</i>	<i>very</i>
adjective	78%	81%
adverb	10%	10%
verb	7%	2%
determiner	5%	7%

As one can see, *too* and *very* are primarily associated with adjectives that represent 78 % and 81% of their collocations respectively. However, the right tokens of *too* and *very* are not always the units modified by the adverb. Sometimes, the adverb modifies the second token in its right contexts, not the first one. Nonetheless, when the right context is an adjective or a verb, it is the modified one by the adverb without exception: when a determiner or an adverb appears in the right contexts of *too* and *very*, it is not always a modified one, but the unit following the first one can be the unit modified by the adverbs. Therefore, we focus in this study only on the adjectival collocations of *too* and *very* to investigate the evaluation types of these adverbs.

For more accurate results, the first 100 collocations were extracted for every adverb. Also, collocations with less than 3 occurrences were excluded. Because of space limitations the table below displays only the first 50 occurrences.



**Table 2.** *50 strongest collocations with our selected adverbs in the BNC*

<i>very</i>	frequency	<i>too</i>	Freq.
Difficult	2286	<b>Far</b>	3275
indeed	1939	<b>Late</b>	2319
Nice	1669	<b>Bad</b>	787
Thank	1654	<b>Busy</b>	558
Quickly	970	<b>Fast</b>	508
carefully	601	<b>Expensive</b>	388
pleased	593	<b>Tired</b>	279
careful	503	<b>Weak</b>	246
attractive	416	<b>Frightened</b>	122
helpful	394	<b>Scared</b>	95
sad	327	<b>Painful</b>	92
keen	326	<b>Embarrassed</b>	81
clever	322	<b>Costly</b>	77
pleasant	314	<b>Risky</b>	69
grateful	286	<b>Shy</b>	63
proud	273	<b>Lazy</b>	62
fond	219	<b>Preoccupied</b>	59
disappointed	175	<b>loudly</b>	52
nicely	131	<b>Polite</b>	52
seldom	121	<b>Simplistic</b>	48
enjoyable	107	<b>Restrictive</b>	45
poorly	89	<b>Little</b>	37
supportive	77	<b>Cumbersome</b>	35
frustrating	71	<b>Lenient</b>	34
talented	71	<b>Sexy</b>	33
distressed	64	<b>Inexperienced</b>	32
amusing	63	<b>Hasty</b>	27
rewarding	57	<b>Complacent</b>	22
distressing	45	<b>dwel</b>	22
tiring	43	<b>Greedy</b>	22
cheaply	42	<b>Fussy</b>	21
good-looking	41	<b>Inflexible</b>	20
windy	37	<b>Harshly</b>	19
time-consuming	32	<b>hassle</b>	17
tasty	31	<b>Pessimistic</b>	16
knowledgeable	28	<b>Timid</b>	15
flattering	27	<b>Juicy</b>	15
stressful	27	<b>Bulky</b>	15
considerate	21	<b>oblige</b>	14
astute	19	<b>Engrossed</b>	13
creditable	18	<b>Lax</b>	12
possessive	17	<b>Fanciful</b>	11
gratifying	16	<b>Glib</b>	11
sketchy	15	<b>One-sided</b>	11

likeable	15	<b>Intrusive</b>	10
obliging	15	<b>Pushy</b>	10
low-key	15	<b>Far-fetched</b>	10
sociable	15	<b>Imprecise</b>	10
anthesis	15	<b>Unwieldy</b>	9
fine-grained	14	<b>Uptight</b>	9
undemanding	14	<b>Sanguine</b>	9
talkative	13	<b>Single-minded</b>	9
commendable	13	<b>Inquisitive</b>	8
approachable	12	<b>Obtrusive</b>	8
touchy	11	<b>Easy-going</b>	8
choosy	10	<b>cocky</b>	8

### *Polarity Types*

In an attempt to investigate how the adverbs analyzed in this study contribute to the polarity of meaning, we classified their collocations as being negative, positive or neutral. The results were as follows:

**Table 5.** *Polarity of the collocations of too and very*

	positive	negative	neutral
too	12%	84%	4%
very	72%	20%	8%

In the case of *very*, the 72 % of the adjectives collocated with *very* are positive adjectives such as *nice*, *pleased*, or *careful*, while only 20 % are negative adjectives such as *sad*, *disappointed*, or *time-consuming*.

*Too*, on the contrary, is primarily associated with negative adjectives (84 %), examples of collocations are *bad*, *tired*, or *expensive*. Only 12 % of the collocations are positive adjectives such as *fast* or *polite*. In other words, the adjectives of negative polarity co-occur with *too* 7 times more frequently than those of positive polarity. On the contrary, the adjectives of positive polarity co-occur with *very* 2.6 times more frequently than those of negative polarity. This is the reason why, the adverb *too* is usually deemed to contribute to a negative evaluation. We notice also that the adverbs *very* and *too* collocate with polarity neutral adjectives, which constitutes respectively 8% and 4% of their collocations. That is, those adjectives that are by virtue neither positive nor negative, and only the context decides their polarity.

In the next sub-sections, we investigate how the adverbs *too* and *very* affect the original sentiment polarity (positive, negative and neutral) of the adjectives they modify. We also compare the two adverbs along the way.

### Too and very + neutral adjectives

*Too* and *very* are both amplifiers (Quirk et al 1985), which means they are both used to express degrees of increasing intensification upwards from an assumed norm. Nonetheless, the two adverbs are dissimilar. While *very* increases the intensity of the degree for the predicates, the adverb *too* not only

increases the intensity of the degree, but under some conditions, it can add a negative evaluation. Consider:

(1a) The room is very large

(1b) The room is too large

In both of the above examples, the adverbs *very* and *too* amplify the intensity of degree for the adjective *large*. Nevertheless, only (1b) reveals a negative evaluation concerning the fact that *The room is large*, while (1a) shows a rather neutral opinion in regards to the polarity such as positive or negative. This negative opinion expressed by *too* in (1b) is often neglected among scholars in linguistics, because it is different from ‘negation adverbs’, which formally introduce a negation on the positive sentences, or ‘the adverbs known as NPIs’, which formally require negative sentences: it simply adds a negative opinion which is not a semantic negation, but a pragmatic negative evaluation.

#### Too and very + positive and negative adjectives

As mentioned before, the adverb *too* is usually considered as a negative intensifier regarding the degree of modified predicates, in opinion mining research as well as in linguistics literature. However, we demonstrate in this section that the adverb *too* is quite similar to the adverb *very* in that they rather play the role of a booster in terms of Quirk et al. (1985), than a polarity reversing marker in terms of Kennedy & Inkpen (2006), especially when associated with negative and positive which constitute together more than 90 % of the adverbs’ collocations in the BNC. Consider the following sentence:

(2) You are smart.

The above example comprises the adjectival predicate *smart*. Due to the semantic property of this adjective, the evaluation of the entity occurring in the subject position is positive. When the adverbs *too* and *very* are introduced to modify the adjective, they simply emphasize the original semantic property of the adjective as shown in the following examples:

(3) a. You are very smart.

(3) b. You are too smart.

In both sentences, the positive property of the adjective *smart* receives a stressed positive evaluation by means of the introduction of the adverbs *too* and *very*. Nevertheless, the two adverbs differs in terms of the degree of intensity in that *too* projects a stronger intensity than *very*. Consider the following example:

(4) a. The pay is very good for me= I am satisfied with it.

(4) b. The pay is too good for me = it is more than what I qualify for.

The evaluation of the entity in the subject position (*the pay*) is clearly more positive in (4.b) than in (4.a). This is due to the higher intensifying property the adverb *too* has over the adverb *very*.

Similarly, the two adverbs *too* and *very* further intensify the negative semantic property of negative adjectives. Consider:

- (5) a. The new computer is very expensive.
- (5) b. The new computer is too expensive.

By all accounts, any entity modified by the adjective *expensive* would be assessed to have a negative sentiment polarity due to the negative semantic property inherited in the adjective *expensive*.

To sum up, the adverbs *too* and *very* do not change the sentiment polarity expressed by positive and negative adjectives, but simply is just emphasizing their original meaning. Because positive and negative adjectives constitute more than 90% of the collocations of *too* and *very*, we believe these adverbs primarily works as meaning boosters at the expense of a very shallow use of *too* with neutral adjectives (4%), which has a subjective orientation rather negatively oriented.

#### Corpus Types

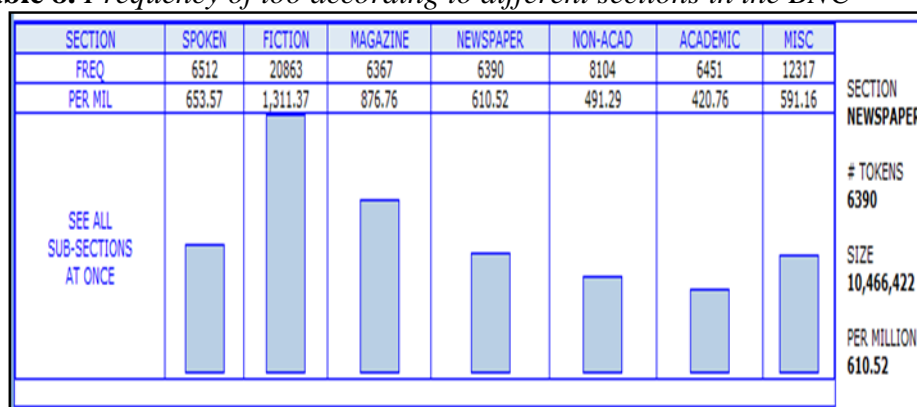
After conducting a frequency search using the chart search option enabled by the BNC, the results we obtained are as follows:

**Table 7.** Frequency of *very* according to different sections in the BNC

SECTION	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	NON-ACAD	ACADEMIC	MISC	
FREQ	25032	22021	9265	8899	15010	14207	25138	
PER MIL	2,512.33	1,384.16	1,275.82	850.24	909.96	926.64	1,206.52	SECTION SPOKEN
								# TOKENS 25032
								SIZE 9,963,663
								PER MILLION 2,512.33

SEE ALL SUB-SECTIONS AT ONCE

**Table 8.** Frequency of *too* according to different sections in the BNC



The two charts above allow us to see clearly the difference between the two adverbs in terms of the scope of their usage. The majority of the occurrences of the adverb *very* were found in the spoken section of the BNC followed by the fiction corpus type. On the contrary, the adverb *too* was highly used in the fiction, followed by magazine, and then the spoken corpus type comes in the third place. This suggests that there is a difference of formality degree between both adverbs. Namely, *very* appears in less formal contexts than *too*.

### Conclusion

The adverbs *too* and *very* are generally known as intensifying a given meaning of modified words, notably adjectives and adverbs. However, based on the empirical examination of the British National Corpus (BNC), we assumed that the prominent sentiment polarity type of the words collocated with *very* is dissimilar to that of the words collocated with *too*. In most previous works, it is asserted that it is due to the particular semantic function of the adverb *too*. Nevertheless, we demonstrated that this interpretation does not result from the special function of *too*, but from the semantic orientation itself of the words collocated with these adverbs.

The experiment we performed in this study reveals that the adjectives of negative polarity co-occur with *too* 7 times more frequently than those of positive polarity and. On the contrary, the adjectives of positive polarity co-occur with *very* 2.6 times more frequently than those of negative polarity. The distribution of part-of-speech in the position collocated with these adverbs, the proportion of words with positive/negative/neutral polarity and the corpus types such as formal/written texts or informal/spoken texts clearly deplore Biber et al. (1999) assertion that adverbs like *very* and *too* are used interchangeably.

## References

- British National Corpus website: <http://corpus.byu.edu/bnc/>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Firth, J.R. (1957). 'A synopsis of linguistic theory 1930-1955' in F.R. Palmer (ed.) *Selected Papers of J.R. Firth 1952-1959*, Longman.
- Graeme, K. (2003). 'Amplifier Collocations in the British National Corpus: Implications for English Language Teaching'. *TESOL Quarterly* Vol. 37, No. 3, pp. 467-487.
- Greenbaum, S. (1974). 'Some verb-intensifier collocations in American and British English'. *American Speech*, 49, 79-89.
- Kennedy, A., Inkpen, D., 2006. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. *Computational Intelligence*. Vol. 22. 110-125.
- Mackin, R. (1978). 'On collocations: Words shall be known by the company they keep'. In P. Strevens (Ed.), *In honour of A. S. Hornby* (pp. 149-165). Oxford: Oxford University Press.
- McCarthy, M. & O'Dell, F. (2005). *English collocations in use*. Cambridge: Cambridge University Press.
- Polanyi, L. and Zaenen, A. (2004). Contextual valence shifters. In Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text, pages 106-111.
- Quirk, Randolph, S. Greenbaum, G. Leech, and J. Svartvik. (1985). *A comprehensive grammar of the English language*. London: Longman
- Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sun, Y.C., & Wang, L. Y. (2003). 'Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty'. *Computer Assisted Language Learning*, 16, 83-94.
- Wouden, T. V. D. (1997). *Negative contexts: Collocation, polarity and multiple negation*. London: Routledge.
- Xiao, R., & McEnery, T. (2006). 'Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective'. *Applied Linguistics*, 27, 103-129.