Literally does not always mean *literally*: a corpus-based diachronic study on *literally* as an intensifier

Semi Park

(Seoul National University)

Park, Semi. 2016. Literally does not always mean *literally*: a corpus-based diachronic study on *literally* as an intensifier. *SNU Working Papers in English Linguistics and Language 14, 124-142.* This paper explores the adverb *literally*'s linguistic state over the past century. Nine-hundred tokens of *literally* are collected from the Corpus of Historical American English and a social network service, Twitter. Three decades (the 1900s, 1950s, and 2000s) of data are examined as well as the data of 2016. The empirical data is investigated by *literally*'s seven pragmatic functions and syntactic patterns. The results of data analysis are not consistent, but the overall results support the idea that *literally* is going through grammaticalization, though it is far from a complete internsifier. **(Seoul National University)**

1. Introduction

The second definition of the adverb, *literally*, was officially added to the major English dictionaries such as *Merriam-Webster, a dictionary of American English*, *Longman Dictionary of Contemporary English*, and *Oxford Dictionary of English*, in around 2013. The original definition of *literally* used to refer a particular fashion, as the word denotes, 'in a literal way,' but the additional definition of 'informal' *literally* is used in an exaggerated way to emphasize a literal or non-literal statement or description. Since *literally* became available to mean 'figuratively' as well, two contradictory senses of *literally* have drawn many people's attention for the last decade. A lot of people have been arguing about the 'proper' use of *literally* online, and the dispute seemed to reach its peak as the major English dictionaries added the second definition of *literally*. Despite the prevalence of debates about the correct use of '*literally*,'

surprisingly there had been a little linguistic research conducted on it. As a passionate linguist, I am interested in inspecting *literally*'s linguistic status, not necessarily in determining the 'correct' usage.

Israel (2002) firstly examined the *literally,* focusing on its linguistic history and semantic change. First of all, by looking into the concordance lines from the Oxford English Dictionary, he found out that the new use of *literally* did not show up in one day; the new role of *literally* as a metalinguistic operator, not denoting a manner, started to appear since 1687 as in 'My daily bread is litt'rally implor'd [Dryden, 1687],' 'Every day with me is *literally* another yesterday for it is exactly the same [Pope, 1708],' 'He had the singular fate of dying *literally* of hunger [Hume, 1762].' Moreover, by examining the modern uses of the word *literally*, he argued that *literally* is in the progress of semantic change or grammaticalization [1]   involving both processes of pragmatic strengthening[2] and semantic bleaching[3]. Specifically, he argued that it goes through a typical development of a large class of intensifiers. Quirk et al. (1985) refers to all degree adverbs as intensifiers, which can be divided into amplifiers (intensifiers that scale upwards, e.g., He is a *really* smart student.) and downtoners (intensifiers that scale downwards, e.g., He is *somewhat* crazy.), but most researchers working on intensifier grammaticalization adopt only amplifiers as intensifiers (King, 2016).

The adverb *literally* appears to be evolving into an intensifier undergoing the same semantic extension, so-called 'modal-to-intensifier shift' (Partington, 1993), which intensifiers like *very, really*, *truly,* and *genuinely* had also gone through (Bolinger, 1972; Klein, 1998); however, Israel (2002) argues that *literally* is not yet a complete intensifier. It is

---

[1] The process by which a word gradually shifts from a lexical to a grammatical function (Hopper & Traugott, 2003). Grammaticalization changes modal adverbs to words that intensify an utterance without changing the meaning (Ito & Tagliamonte, 2003).
[2] Pragmatic strengthening refers to the phenomenon where speakers lose consciousness of the original meaning of a word by associating with a new idiomatic meaning of the word (Cole, 1975).
[3] Semantic bleaching refers to the loss of all (or most) lexical content of an entity while only its grammatical content is retained (Heine, 1993).

because even though the current *literally* can amplify an expression, it cannot force a scalar reading to an expression unlike other intensifiers such as *truly*, *very*, or *really*. True intensifiers can even force a scalar construal when nothing seems to be valid; for example, in *really dead* or *very pregnant,* death and pregnancy can be viewed as scalar properties. Yet, *literally dead* or *literally pregnant* do not give death and pregnancy scalar properties.

About a decade later, Calhoun (2013) carried out more or less empirical research in order to inspect the syntactic and pragmatic functions of modern uses of *literally*, as well as to investigate speakers' general attitudes toward the new uses of the word, and to organize what are the common arguments criticizing or defending the use *literally* as an intensifier and so on. She answered those research questions by analyzing 1) 125 tokens of *literally* which were selected at random from public media, primarily social media, 2) popular discourses from the entertainment websites (e.g. Buzzfeed, College Humor), blogs, and online commentary, 3) an online survey completed by 270 university students 18 to 25 years of age, 4) a focus group interview with four university students.

From the dataset, she found that the tokens of *literally* primarily modified the types of phrases expected for an adverb, i.e. verb phrases and adjective phrases. A few tokens of *literally* modifying noun phrases also prepositional phrases appeared as well. When it comes to pragmatic patterns, she argued that *literally* has six major functions; one of them was a function of an intensifier. In her dataset, which consisted of the number of 125 tokens, 61% of tokens were strengthening speakers' utterances. Based on the result, she also claimed that *literally* is also in the process of grammaticalization in line with Israel (2002).

Another finding was about people's perception of *literally* as an intensifier. Nearly all (96%) of the respondents of the online survey, who were university students, acknowledge that *literally* can function as an intensifier. Although university students are the ones who use the word

as an intensifier the most frequently, the majority of attitudes were not positive ironically. Overarching attitude towards intensifying *literally* was that it should only be used in informal contexts rather than in formal contexts such as professional settings (e.g., job interviews) or in academic writings. In sum, many people know that *literally* can function in a figurative sense, but they still treat it like slang, which has limited registers.

1.1   Current study

The current study focuses on the diachronic investigation of the syntactic and pragmatic patterns of *literally* using American English corpora. Israel (2002) and Calhoun (2013) argued that *literally* is in the process of grammaticalization into an intensifier, but they analyzed the tokens of *literally* mainly in a synchronic way. Based on the fact that the frequency of *literally* increased about 64 percent in last fifty years, this study was designed to inspect empirically whether the increase resulted from the rising of the new meaning of *literally* as an intensifier, or it was a coincidence.

2.   Research Method
2.1   Data Collection and Analysis

For the current research, 750 data points were collected from the Corpus of Historical American English (COHA), and the number of 150 data was randomly collected from an online social networking service, Twitter. In total, 900 data points were examined by the author and one native American English speaker in case of need. All statistical analysis was completed using Microsoft Excel 2016.

2.1.1   The Corpus of Historical American English

The Corpus of Historical American English has more than 400 million words of text in more than 100,000 individual documents, created between 1810 and 2009. The corpus is composed of four subcorpora, which are Fiction, Magazine, Newspaper, and Non-fiction, and it is balanced by the genre across the decades; for example, Fiction accounts for about 50% of the total in each decade (1810s-2000s). The corpus is also balanced across decades for sub-genres and domains as well.

Literally appeared in the corpus 14.75 times per million in the 1900s, 17.80 times per million in the 1950s, and 27.83 per million in the 2000s, respectively. In order to investigate the trend of *literally*'s usage in a diachronic manner, three periods (the 1900s, 1950s, and 2000s) of data were collected from Fiction and Magazine subcorpora of the COHA. Two subcorpora were selected because they are more informal than Newspaper and Non-fiction, and their data averagely take 75% of the corpus. From the Fiction subcorpus, 150 data points were collected from each decade, in total 450, and from the Magazine subcorpus, 100 data[4] points were collected from each decade, in total 300, as well. The reason for using that specific number of data points was that 150 was the maximum number of the Fiction data entries available in the 1950s, and 100 was the maximum number of the Magazine data entries I could get in the 1900s.

2.1.2   Twitter

Since the COHA only contains data until 2009, a social network service, Twitter, was used to collect the latest usage of *literally*. Twitter as a web-based corpus is a rich source of linguistic information. As of December,

---

[4]   There were only 93 concordance lines between 1900 and 1909 in Magazine subcorpus, so I had to add seven more examples from 1899's Magazine corpus to my dataset to examine 100 concordance lines.

13th, 2016 6:42 P.M., 7,441[5] tweets are sent in 1 second on Twitter, which corresponds to over 446,460 tweets sent per minute, about 645 million tweets per day and around 235 billion tweets per year. It is hard to get demographic information of the users, but it is very useful to get recent, informal and colloquial linguistic data. The data for this study was collected from November 21st, 2016 through November 28th, 2016, using Twitter Advanced Search[6]. I put '*literally*' into the '*all these words*' slot and chose '*English (English)*' in the '*written-in*' dropdown selection in the Twitter Advanced Search web page. The number of 150 data that are comprehensible without needing further context was collected.

2.1.3   Analysis

First of all, in order to inspect if the new definition of *literally* contributed to the increased frequency of the word in the corpus, the collected data was classified based on the *literally's* function in each concordance line. Data analysis was based on the six major functions of *literally* suggested by Calhoun (2013), but there was one modification to her criteria. After all the data was collected, I examined 100 examples from the dataset as a pilot examination to see if Calhoun (2013)'s classification is enough to categorize every usage of *literally*, and as a result of the preliminary investigation, one more function of *literally* was added. The additional function of literally (1b) is to denote the literal sense of linguistic objects 'word by word,' as in examples follow: '*" Is that message to be conveyed literally? " # " Yes! "*'; '*Don't take everything so literally.*'
Calhoun (2013) revealed six primary functions of *literally* using functions of *literally* in previous research as guidelines (Israel, 2002; Liberman, 2011; Lukes, 2011). The new list of functions of *literally*

---

[5]   The website helps you to count the number of Tweets per second. (http://www.internetlivestats.com/one-second/#tweets-band)
[6] You can list up people's Tweets containing specific words or phrases using Twitter Advance Search. (https://twitter.com/search-advanced?lang=eng)

below is, therefore, based on those research papers. The majority of the seven functions are not mutually exclusive.

1a. To indicate translation, 'word for word'
   (1) …however, that the English name is derived from the Dutch name kruisbes, *literally*, " cross-berry. " [COHA:1951:MAG]

1b. To denote the literal sense of linguistic objects 'word by word'
   (2) " I grew up around people who took the Bible *literally*, and still do. " [COHA:2000:MAG]

2.   To arbitrate between two literal expressions
   (3) Until your baby becomes more comfortable with your new schedule, he may *literally* " milk " you for as much as he can during the night. [COHA:2004:MAG]

3.   To avoid more awkward or less-fitting word choice
   (4) God is *literally* my COVERING The countless times He has protected me from situations and potential hurt because of His love for me [Twitter]

4.   To invert a figurative expression, 'in a literal sense'
   (5) Then, in a flash, he dropped flat on the ground, and *literally* licked my shoes. [COHA:1901:MAG]

5.   To highlight apt or clever word choice
   (6) If you wanted to go upstairs here, you *literally* had to go up the stairs. [COHA:2008:FIC]

6.   To strengthen the force of an utterance by marking speaker commitment
   a.   Literal expression

1) General utterance
(7) do u ever crave someone's presence so much you would
   literally be happy just be sitting beside them silent [Twitter]
2) Number or measurement
(8) FINALS R *LITERALLY* IN 3 WEEKS HOLD MY HAND AND
   MY GRADES [Twitter]

b. Non-literal expression
1) Figurative: simile or metaphor
(9) She was so perfectly groomed that she looked as though her
   clothes were a mould into which she had *literally* been poured.
   [COHA:1909:FIC]
2) Hyperbolic
(10) She knows everything, *literally* everything... in her field, that
   is. [COHA:1952:FIC]

c. Humor (literal or non-literal expression)
(11) … in particular about the way his chest hair the color of
   champagne seemed *literally* to bubble out of the collar of his
   sweater. [COHA:2000:FIC]

d. Insult (literal or non-literal expression)
(12) Wish we could play Arsenal, Chelsea and City etc each week
   instead of teams who *literally* come here for the one point. Hate
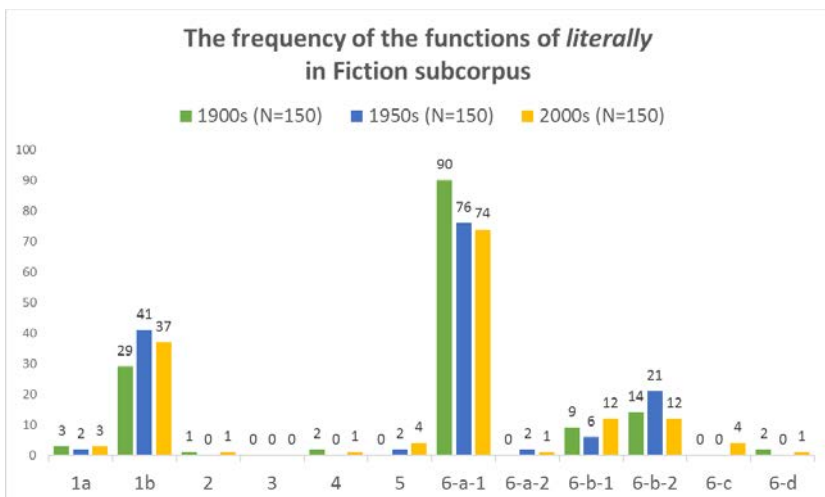   games like this. [Twitter]

Second, to see if there had been changes in the word's syntactic patterns, phrases which were modified by *literally* were examined in each token. Partington (1993) argued that the more grammaticalized a modal adverb is, the more widely it collocates; therefore, probing the syntactic patterns of *literally* diachronically can suggest further evidence of the degree of its grammaticalization.

3.   Results

3.1   *Literally*'s pragmatic functions

In general, the majority of tokens were used as intensifier to strengthen the force of an utterance by marking speaker commitment (function 6) in the collected data. Considering that the functions from 1a through 5 have specific senses, and the functions from 6a through 6d simply work as intensifier without adding no content to a sentence, the definition of *literally* can be divided into two big categories: meaningful *literally* versus grammatical *literally* or intensifier. *Literally* as an intensifier was used most frequently on Twitter, followed by Magazine and Fiction. Figure 1, 2, and 3 represent the frequency of the functions of *literally* in Fiction subcorpus, Magazine subcorpus, and Twitter data, respectively.
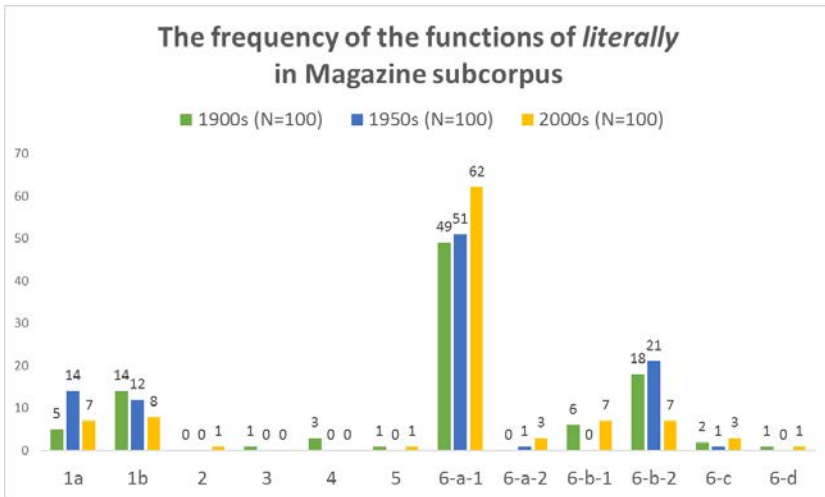
Figure 1. The frequency of the functions of *literally* in Fiction subcorpus in the COHA

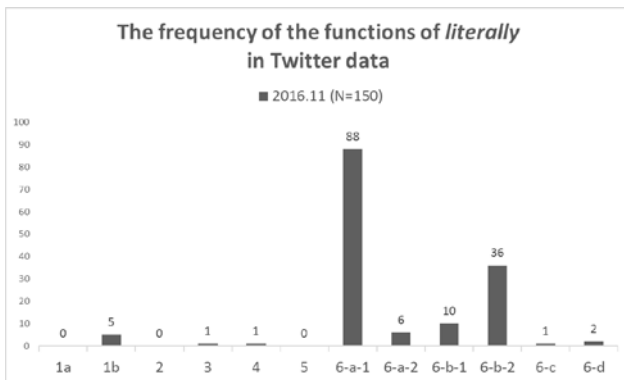In Fiction subcorpus data shown above, meaningful *literally* (1a through

5) occurs on average 28 percent, and grammatical *literally* (6-a-1 through 6-d) occurs on average 72 percent for last 100 years. Tokens of function 3 were not observed at all, and tokens of function 1a, 2,4, 5, 6-a-2, 6-c and 6-d were rarely observed. Only the function 6-a-1, which intensifies a general utterance, experienced a decrease in usage over the hundred-year period.

Figure 2. The frequency of the functions of *literally* in Magazine subcorpus in the COHA

In Magazine subcorpus data, meaningful *literally* (1a through 5) occurs on average 22.33 percent, and grammatical *literally* (6-a-1 through 6-d) occurs on average 77.67 percent over the last hundred years. Tokens of function 6-b-1 were occasionally observed, but 2,3, 4, 5, 6-a-2, 6-c and 6-d were rarely observed – the frequency was less than ten times in hundred-year of period . Unlike in the Fiction data, the function 6-a-1 is on increase, and the function 1b is on decrease over the 100-year period in the Magazine subcorpus data.
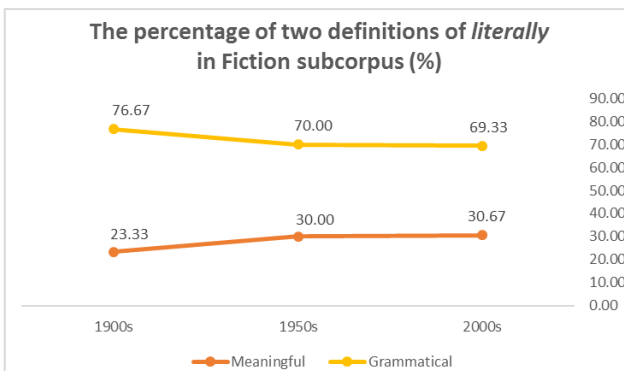
Figure 3. The frequency of the functions of literally in Twitter data

The frequency of the functions of *literally* in Twitter data

In Twitter data, meaningful *literally* has an occurrence of only 4.67 percent, and grammatical *literally* has an occurrence of 95.33 percent. Especially, function 6-a-1 is overwhelmingly used, followed by the function 6-b-2. Tokens of function 1a, 2 and 5 were not observed at all, and tokens of function 3, 4, 6-c and 6-d were rarely found.
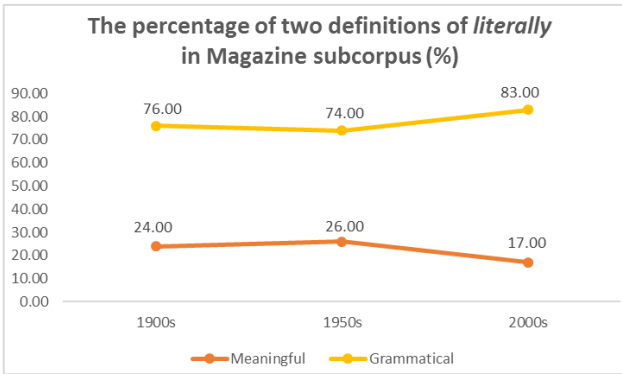
Figure 4 and 5 denote the percentage of two definitions of *literally* for each explored decade in Fiction and Magazine subcorpus respectively. As the rate represents, grammatical *literally* takes the majority in both Fiction and Magazine data.

Figure 4. The percentage of two definitions of literally in Fiction subcorpus

The percentage of two definitions of *literally* in Fiction subcorpus (%)

In Fiction subcorpus data, there is a tendency of increase for meaningful *literally* and a tendency of decrease for grammatical *literally*.

Figure 5. The percentage of two definitions of literally in Magazine subcorpus

In the Magazine subcorpus data, on the contrary, there is a tendency of decrease for meaningful *literally* and a trend of increase for grammatical *literally* over 50 years.

3.2   *Literally's* syntactic patterns

Table 1 shows the average proportion of modified phrases. The percentages of three decades were calculated by averaging the ratio of two subcorpus data of each decade.

Table 1. The average percentage of modified phrases

	1900s	**1950s**	**2000s**	**2016**
NP, PP, ADVP	21.17 %	26.67 %	24.00 %	23.33 %
VP, AP	78.83 %	73.33 %	76.00 %	76.67 %

Note: NP = noun phrase, PP = prepositional phrase, ADVP = adverbial phrase, VP = verb phrase, AP = adjective phrase
Overall, there was no dramatic change in the word's syntactic patterns in

116 years, despite the slight differences between analysis data. *Literally* was primarily used to modify verb phrases as in (13) and adjective phrases as in (14), but it was also sometimes used to modify noun phrases (15), prepositional phrases (16), and adverbial phrases (17). The figures of separate results of each corpus are represented in Appendix section.

(13) …But it's kind of tough when you *literally* have to scoot over… [COHA:2000:FIC]

(14) This is all very hard. Those specks almost made me *literally* sick. [COHA:2000:FIC]

(15) … he ran a program that was quite *literally* a whorehouse;… [COHA:2000:MAG]

(16) She swam every day in the sea *literally* until the day she died. [COHA:2000:MAG]

(17) …, which make Jack *literally* as good as his master…. [COHA:1907:MAG]

4.   Discussion

When it comes to the emergence of a new definition of *literally*, a few researchers argued that it derived from *literally's* grammaticalization. The purpose of the current study was to empirically investigate if the increased frequency of the modal adverb *literally* can actually be explained by its grammaticalization into an intensifier. As the increase of general frequency and extension of collocations are considered as the by-product of grammaticalization, diachronic changes of syntactic patterns and increasing proportions of intensifying *literally* were expected to represent. The Corpus of Historical American English and Twitter were used to collect diachronic data, and a total of 900 tokens of *literally* were analyzed.

Two subcorpora, Magazine and Fiction, from the COHA were chosen because those two corpora take around 75 % of the entire data, and they

are more informal than the other two subcorpora, Newspaper and Non-fiction. As linguistic change tends to occur in informal settings first, Magazine and Fiction subcorpora were considered to have more appropriate data to examine the state of *literally*. Since the latest COHA corpus data was from 2009, a social networking service, Twitter, was used as a kind of web-based corpus to inspect the most recent and informal use of *literally*.

The syntactic and pragmatic analysis was conducted on all collected tokens of *literally*. The results of each dataset were variant. First of all, the proportion of *literally* as an intensifier (i.e., grammatical *literally*) has already taken the majority of its use since the 1900s or even earlier. The percentage of intensifying *literally* was on the increase only in the dataset from the Magazine subcorpus. As for the tokens from the Fiction subcorpus, on the contrary, the proportion of *literally* as an intensifier was on the decrease. The inconsistent results are confusing because they keep us from answering the research question – whether the increased frequency of *literally* in the entire corpus is due to the growing use of intensifying *literally*. Nevertheless, the result of Twitter dataset analysis denotes that *literally* can be overwhelmingly used as an intensifier in the case of very informal settings.

The fact that the results are varying for different dataset signifies two things. First, a target feature's dichromic history of change and its current state cannot be solely determined by dataset of a small number of registers and sources. Second, this means a corpus is not almighty. Of course, corpus linguistic method allows us to look into a large number of authentic language data produced by a variety of speakers or writers and helps to confirm or contest hypotheses about the specific features based on empirical evidence. However, as different corpora contain different data, and different data can cause different results, some features may or cannot appear in the data depending on the chosen corpora. For example, if I looked into a different Magazine corpus, which is composed of different magazines, I would have gotten different aspects of *literally*.

This signifies that the results of corpus-based data analysis cannot always reflect the absolute truth of the linguistic object. In addition to this, corpus data can also have wrongly transcribed tokens[7], though it is very rare, which can mislead researchers; hence, researchers should be careful using corpus-based method for research.

When it comes to syntactic patterns of *literally*, there was not a consistent trend over time. Similar to the historical proportions of grammatical *literally*¸ the modal adverb *literally's* collocation was already widespread in the 1900s; it has been modifying other phrases as well as verb phrases and adjective phrases. Meanwhile, there was an interesting characteristic of *literally* in the data. Though it is known that modal adverbs increase their range of collocation during the process of grammaticalization (Iso & Tagliamonte, 2003), our *literally* is a little more special than other modal adverbs. As *literally* has a unique function of 'word-for-word' translation, it seems inevitably likely to be collocated with noun phrases from the beginning. (e.g., "An amparo (*literally*: protection) is a Mexican legal confection with elements of habeas corpus…"[COHA:1950:MAG]) In other words, for *literally*, frequent collocations with noun phrases cannot be simply regarded as an evidence of *literally*'s grammaticalization.

In summary, the data from this study is not enough to argue that the scope

---

[7] In the Magazine subcorpus data of 1907, there is one error in the given text transcribed from *The Atlantic Monthly*. The transcribed text in the corpus is:

> 'He finds much to revolt him in the manners of the people at table and elsewhere, but is more than half won over by the spontaneous sympathy which makes a shopwoman say, "How do you do."" to you as you enter.M. Huret makes the shrewd remark that the American independence of manner is often due less to any high moral sentiment about the equality of man than to actual circumstances of condition and origin, which make Jack **literally** as good as his **toster**.'

The word *toster* was tricky because it actually gave me a hard time to decide the function of *literally*. So I looked into the magazine's archive online and figured out that the word *toster* in the last sentence was actually *master* in the original magazine. (see: https://www.unz.org/Pub/AtlanticMonthly-1907oct-00553 / p.557 / Some Recent Books on the United States   by James F. Muirhead Vues d'Amerique, by Paul Adam)

of *literally* are getting more and more widespread, and the frequency of *literally* increases because of its popularized intensifying function, as was hypothesized. The periods of data were not long enough, and the kinds of registers were not various enough either. Nonetheless, the current study can support the previous argumentation of Israel (2002) and Calhoun (2013) that *literally* is being grammaticalized at present. That is possible because the diachronic data shows that *literally* has taken various syntactic patterns for the last 100 years, and the word has already been mainly used as an intensifier at least in informal settings. Besides, Twitter data denoted that *literally* can be overwhelmingly used as an intensifier in the case of very informal settings.

5. Conclusion

It seems that *literally* is going through its grammaticalization into a better intensifier. As other modal adverbs, which fully function as intensifiers today, such as very, really, completely, etc., *literally* has widespread collocations, and its general frequency has increased. However, whether *literally* will successfully become a complete intensifier or not is in question. It is not merely because people have negative attitudes to the intensifying *literally*. Rather, it is because its original meaning is too special to be lost. During the grammaticalization, words start to lose their lexical content so that they can have another function (Partington, 1993). In the case of intensifiers, original senses of words start to be reduced to have an amplifying function, and through pragmatic strengthening over time they are finally deprived of their original lexical meaning, but convey simple intensification only. For example, the original meaning of an intensifier *very* was 'true' or 'real,' (Ito & Tagliamonte, 2003) but not many people know that its original meaning was 'truth.' In the case of *very*, however, it could have been possible for it to lose its original meaning completely since there existed other substitutable words of truth.

The original meaning and function of *literally* is very specific, so *literally* might not become a perfect intensifier unlike *very*. Nobody can predict a language change completely accurately, but as *really* also took about 400 years to be accepted as an intensifier, (Ito & Tagliamonte, 2003) our descendants may see the end of *literally*'s semantic change.

References

Bolinger, D. (1972). Degree words. The Hague: Mouton.

Calhoun, K. (2015). "It is the worst of our time": Youth language, language attitude, and arguments about *literally*. In *Proceedings of the 23rd Annual Symposium about Language and Society-Austin*.

Cole, P. (1975). The synchronic and diachronic status of conversational implicature. *Syntax and semantics, 3,* 257-288.

Israel, M. (2002). Literally speaking. Journal of Pragmatics, 34(4), 423-432. doi:10.1016/s0378-2166(01)00048-0

Ito, R., & Tagliamonte, S. (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers. Language in Society, 32(02). doi:10.1017/s0047404503322055

Klein, H. (1998). Adverbs of degree in Dutch and related languages. Amsterdam: J. Benjamins Pub.

Langacker, R. W. (n.d.). Subjectification, grammaticization, and conceptual archetypes. Subjectification. doi:10.1515/9783110892970.17

Partington, A. (1993). Corpus evidence of language change – the case of the intensifier. *Text and Technology. In Honour of John Sinclair*, 177-192.

Traugott, E. C. (1995). Subjectification in grammaticalisation. *Subjectivity and Subjectivisation*, 31-54.

Appendix

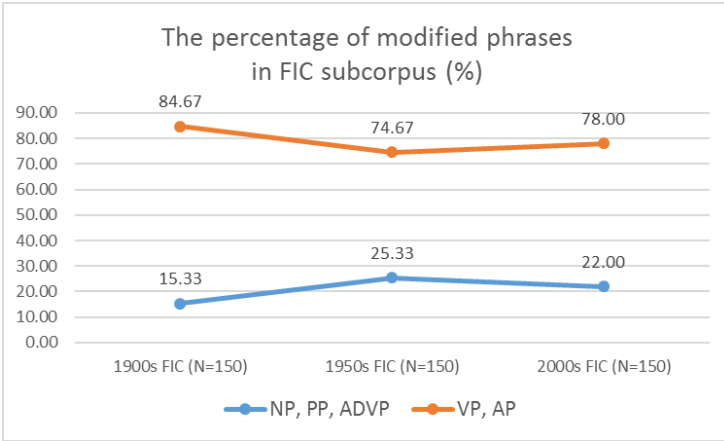Figure 6. The percentage of modified phrases in FIC subcorpus
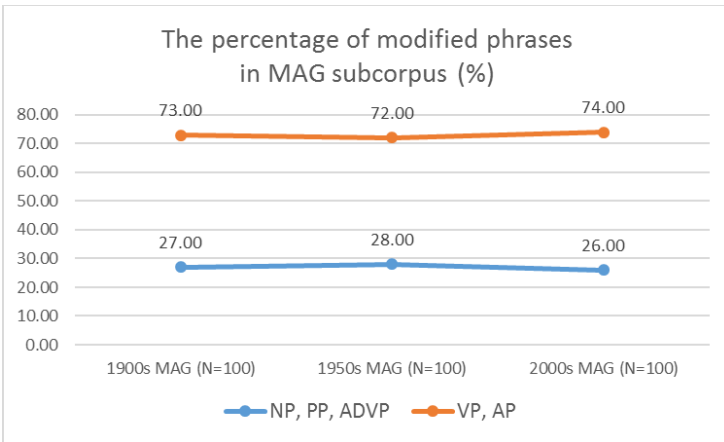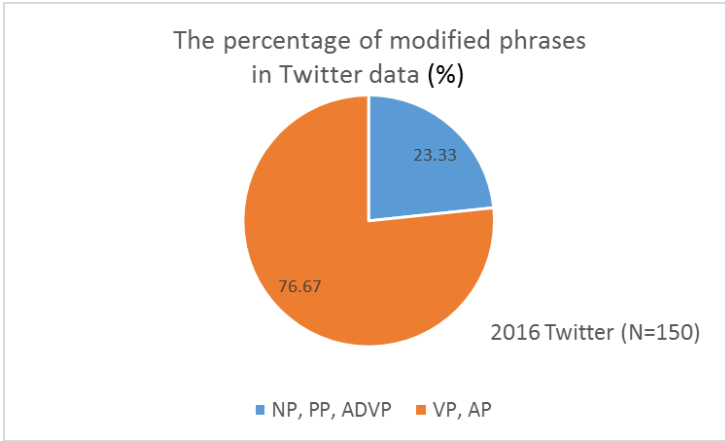
The percentage of modified phrases in FIC subcorpus (%)

Figure 7. The percentage of modified phrases in MAG subcorpus

The percentage of modified phrases in MAG subcorpus (%)

Figure 8. The percentage of modified phrases in Twitter data

The percentage of modified phrases
in Twitter data (%)

23.33

76.67

2016 Twitter (N=150)

■ NP, PP, ADVP   ■ VP, AP

Semi Park
semipark@snu.ac.kr