

Exploring Long-Term Temporal Trends in the Use of Multiword Expressions

Tal Daniel

Ben-Gurion University of the Negev
Beer-Sheva, Israel
dantal@post.bgu.ac.il

Mark Last

Ben-Gurion University of the Negev
Beer-Sheva, Israel
mlast@bgu.ac.il

Abstract

Differentiating between outdated expressions and current expressions is not a trivial task for foreign language learners, and could be beneficial for lexicographers, as they examine expressions. Assuming that the usage of expressions over time can be represented by a time-series of their periodic frequencies over a large lexicographic corpus, we test the hypothesis that there exists an old–new relationship between the time-series of some synonymous expressions, a hint that a later expression has replaced an earlier one. Another hypothesis we test is that Multiword Expressions (MWEs) can be characterized by sparsity & frequency thresholds.

Using a dataset of 1 million English books, we choose MWEs having the most positive or the most negative usage trends from a ready-made list of known MWEs. We identify synonyms of those expressions in a historical thesaurus and visualize the temporal relationships between the resulting expression pairs. Our empirical results indicate that old–new usage relationships do exist between some synonymous expressions, and that new candidate expressions, not found in dictionaries, can be found by analyzing usage trends.

1 Introduction

In this work, we explore Multiword Expressions (MWE) usage over a period of a few hundred years. Specifically, we focus on English MWEs of 2–3 words with long-term decreasing or increasing usage trends that exist in a ready-made list of MWEs. We do not focus on semantic change of these expressions, which is another research field.

From a list of MWEs with statistically significant trends, we try to identify a subset of expressions that have an inverse usage relationship with their near-synonymous expressions, replacing them, or being replaced by them over time.

Another objective of this work is to find potentially new candidate MWEs in a list of collocations that withstand certain sparsity & normalized frequency thresholds and have a statistically significant trend over the years. The normalized frequency threshold represents the minimum number of a collocation mentions, whereas the sparsity threshold represents the minimum number of years, or periods, a collocation is used (not necessarily in consecutive order), making a distinction between real MWEs and temporarily used multiword expressions.

2 Related Work

2.1 Multiword Expressions (MWEs)

Languages contain Multiword expressions (MWEs) that are compounded from a few words (lexemes). MWEs contain various types of expressions such as transparent collocations, fixed phrases, similes, catch phrases, proverbs, quotations, greetings, & phatic phrases (Atkins and Rundell, 2008). They are also used “to enhance fluency and understandability, or mark the register/genre of language use [...]. For example, MWEs can make language more or less informal/colloquial (c.f. *London Underground* vs. *Tube*, and *piss off* vs. *annoy*).” (Baldwin and Kim, 2010) Some MWEs are idiomatic expressions (e.g. *pull one’s leg*), while others “[...] have the singularity of breaching general language rules” (Ramisch, 2013, p2), such as *from now on*, *from time to time*, etc. They may be common names, e.g., *master key*, *vacuum cleaner*, and “sometimes the words [...] are collapsed and

form a single word” (Ramisch, 2013, p2), like *honeymoon*, and *firearm*.

Since MWEs are a mixed set with multiple phenomena, we adopt the broad and practical definition that Ramisch (2013) used, based on Calzolari et al. (2002): “[...] phenomena [that] can be described as a sequence of words that act as a single unit at some level of linguistic analysis” (Ramisch, 2013, p23). This definition emphasizes that MWEs are a single unit, which is especially important for translation, as Ramisch hints.

Several methods exist for finding, or extracting, MWEs from corpora. Often, researchers focus on a single kind of expressions, and length, e.g., Noun-Noun expressions of length two (Al-Haj and Wintner, 2010), or Verb-Noun idiom construction (Fazly et al., 2009). Focusing on a certain kind of expressions can be achieved by crafting a tailored-characterization of these MWEs, creating a model using a machine learning algorithm, and testing it. For example, Tsvetkov & Wintner (2011) suggested a method for any kind of MWEs, by training a system to learn a Bayesian model, based on characteristics such as the number of contexts the expression occurs in, how flexible it is to synonym word replacements, syntactic variability, or whether a translation of the expression appears in another language.

2.2 Trend Detection in Language Corpora

As new expressions become less, or more, frequently used, we can try to track these changes over the years by finding frequency trends. Identifying a trend involves a few tasks, though: One has to identify a statistically significant change in the data over time, to estimate the effect size of that change, while trying to pinpoint the exact time periods of these changes (Gray, 2007).

Buerki (2013) compared three methods for finding “ongoing change” in MWEs within Swiss Text Corpus, which he divided into 5 periods, or data points. He found that the Chi-square test was the most flexible, had an arbitrary cut-off frequency value when stating a statistically significant change in frequency, and could alert of a trend when it occurred in some periods, compared to other methods – not only to a continuous linear increase/decrease. Chi-square outperformed other methods as coefficient of difference (D) by Belica (1996) – the sum of squares of frequencies for each period, or coefficient of variance (CV) ,

which ranks the terms and uses an arbitrary cut-off point, e.g., the top third of the ranked list (Buerki, 2013). When the assumption of normal distribution is unrealistic or when the actual trend is non-linear, Kendall’s τ nonparametric statistic (Gray, 2007) can be used.

2.3 Synonymy

Synonymous expressions can replace each other to convey the same meaning. This claim is not accurate, though, since most synonyms are not semantically identical: “Synonymy, or more precisely near-synonymy, is the study of semantic relations between lexemes or constructions that possess a similar usage” (Glynn, 2010, p2). While Glynn’s Cognitive Linguistics research investigated differences between *annoy*, *bother*, and *hassle*, Kalla (2006) studied differences between three Hebrew words that mean a friend: *yadid*, *rea*, *amit*.

Mahlow & Juska-Bacher (2011) created a German diachronic dictionary by finding variations of pre-selected expressions. Expression variations were found by using patterns and by assigning expressions to types (categories). Juska-Bacher & Mahlow (2012) elaborate more on their semi-automatic method to find structural and semantic changes in German phrasemes (idiomatic MWEs): First, they found candidate phrasemes by looking at nouns with at least 2% frequency, as well as other indicators. Then, they chose select phrasemes, after manually looking into old and contemporary dictionaries. These phrasemes were found in various corpora and manually analysed for changes. Above all, their work emphasizes the importance of manual examination, in addition to corpus-based approaches: “Fully automatic detection of phrasemes is not as yet possible, which is why lexicographers have to manually determine idiomaticity (Rothkegel, 2007)” (Juska-Bacher and Mahlow, 2012, p8).

Dagan & Schler (2013) used a semi-automatic iterative and interactive approach for creating a diachronic Hebrew thesaurus. They tried to automatically find synonym terms for a given list of terms by using second-order distributional similarity. Then they let a lexicographer to either select synonyms, or mark terms for query expansion. Kenter et al. (2015) presented an automatic algorithm that detects vocabulary change for specific input terms in Dutch, across a period of 40 years. They used distributional similarity to find time-

stamped semantic spaces, and used the resulting graph to infer synonymous relationship.

3 Research Methods

3.1 Trend Detection & Analysis

To identify increasing and decreasing trends, we calculated the number of yearly mentions in the Google Syntactic Ngrams corpus for each MWE from the jMWE list. Then, we normalized the frequencies by dividing each yearly frequency by the number of words in the corpus for that year. Finally, we segmented the histograms into 7-year periods, summed-up the normalized frequencies in each period, and smoothed the histograms by using a simple moving average with a sliding window size of 5 periods.

Since we segmented and smoothed the time-series, the assumption of sample independence could not be assumed. Hence, we chose two non-parametric tests for trend existence: Kendall’s τ correlation coefficient and Daniels test for trend. Kendall’s τ correlation coefficient is often used when distributional assumptions of the residuals are violated or when there is a nonlinear association between two variables” (Gray, 2007, p29). The null hypothesis of Kendall’s τ is that there is no trend ($H_0 : \tau = 0$), and the alternative hypothesis is that there is a trend ($H_1 : \tau \neq 0$).

Since the values in a time-series are ordered by time, let G_i be the number of data points after y_i that are greater than y_i . In the same manner, let L_i stand for the number of data points after y_i that are less than y_i . Given this, Kendall’s τ coefficient is calculated as

$$\tau = 2S/n(n-1) \quad (1)$$

where S is the sum of differences between G_i and L_i along the time-series:

$$S = \sum_{i=1}^{n-1} (G_i - L_i) \quad (2)$$

The test statistic z is calculated by

$$z = \frac{\tau}{\sqrt{2(2n+5)/9n(n-1)}} \quad (3)$$

When n is large (e.g., $n > 30$), z has ”approximately normal distribution”, so a p-value can be based on the normal distribution table. For smaller n values, other tables can be used to get a p-value

(Gray, 2007). Daniels test for trend (1950, as mentioned in U.S. Environmental Protection Agency, 1974) uses Spearman’s ρ rank correlation coefficient, which ranks each data point X_i in the time-series as $R(X_i)$. After ranking, ρ is calculated as

$$\rho = \frac{\sum_{i=1}^n [R(X_i) - i]^2}{n(n^2 - 1)} \quad (4)$$

As with the Kendall’s τ correlation test, Daniels test compares Spearman’s ρ to a critical value, set by the sample size n : When $n < 30$, the critical value W_p for a desired p-value is set according to a dedicated table (U.S. Environmental Protection Agency, 1974). When $n \geq 30$, the critical value is calculated using X_p , which is the p quantile of a standard normal distribution:

$$W_p = \frac{X_p}{\sqrt{n-1}} \quad (5)$$

We ordered the list of computed trends by the statistic (Kendall’s τ) and reviewed the top 30 expressions with the highest increasing trend and the 30 expressions with the lowest decreasing trend. The usage trends of these 60 expressions were tested again, using Daniels test for trend. Then, we looked up each expression in Oxford Historical Thesaurus¹, tried to find its synonymous expression, and compared the trends of both expressions to visualize an old–new relationship between them.

3.2 Finding New MWEs

We have tested the hypothesis that new MWEs can be detected in a collocations dataset by certain sparsity and normalized frequency thresholds. Using the Google Syntactic Ngrams corpus and the ready-made list of 65,450 MWEs (Kulkarni and Finlayson, 2011), which is used by the jMWE library for detecting MWEs in text, we set the minimum normalized frequency threshold to that of the least mentioned MWE. In the same manner, we set the threshold of maximum sparsity to the sparsity of the MWE that was mentioned in the corpus across the smallest number of years. Next, we compared three criteria for selecting candidate expressions from Google Syntactic Ngrams (collocations) that are not part of the ready-made MWE list: (1) by their top trend statistic and normalized frequency, (2) by their top normalized frequency

¹<http://www.oed.com>

only, or (3) by their lowest sparsity. For each criterion, we labeled the top k collocations as MWEs or not, according to our understanding, and calculated the *precision@k*. The trend statistic criterion was chosen based on the assumption that emerging MWEs are characterized by a positive usage trend until their common adoption.

The code we used, as well as the results can be found on Github².

4 Experimental Results

4.1 Dataset

We found the Google Books Syntactic-Ngrams dataset³ suitable for our needs (Goldberg and Orwant, 2013), since it is a historical corpus containing data over hundreds of years. Specifically, we explored MWE usage using the 1 Million English subset of the dataset that was constructed from 1 Million English books corpus (Michel et al., 2011) published between 1520 and 2008 and originally contained 101.3 billion words. Each line in the dataset already contains 2–5 n-gram (words) collocations that were found in the 1M English corpus at least 10 times. Each collocation entry specifies its terms, part-of-speech tagging and syntactic dependency labels, total frequency, and a frequency histogram for the years where the n-gram was found. For example, here is how a line from the dataset looks like:

```

employed          more/JJR/dep/2
than/IN/prep/3    em-
ployed/VBN/ccomp/0  12  1855,1
1856,2 1936,2 1941,1 1982,1 1986,1

```

For our research, we only used the “arcs” files of the dataset, which contain trigrams – two content words and optionally their functional markers. Content words are meaningful elements whereas functional-markers “[...] add polarity, modality or definiteness information to the meaningful elements, but do not carry semantic meaning of their own.” (Goldberg and Orwant, 2013, p3). These phrases were checked against jMWE’s predefined MWE list (Kulkarni and Finlayson, 2011), which is described later. Although one can explore files with single-word terms as well, tracking their usage should be problematic as they may be polysemous, i.e. their meaning may vary depending

²<https://github.com/Tal-Daniel/daniel-last-MWEs>

³<http://commondatastorage.googleapis.com/books/syntactic-ngrams/index.html>, Version 20130501.

on context and language changes. We assume that polysemy of multi-word expressions is so rare that it can be ignored. Since the jMWE parser relies on part-of-speech tagging to find MWEs, we did not differentiate collocations by their syntactic dependency, and summed histograms with similar part-of-speech (POS) in the dataset into a single histogram, even though they could have different syntactic dependencies.

In order to bring the words to their stem form before sending the trigrams to jMWE expression detector, we lemmatized the terms with Stanford CoreNLP Toolkit (Manning et al., 2014). In addition, due to the special function underscores (“_”) have in jMWE, we converted them to dashes (“-”). If that was the only character of the token/term, it was ignored. The total counts of the number of tokens in the corpus were taken from the Google Books 1M English Corpus (Google, 2009).

4.2 Usage Analysis of Multiword Expressions

For the Google Syntactic Ngrams dataset, we created expression histograms for the years 1701–2008, since only from 1701 there is more than 1 book per year. As a result, histograms spanned 309 years instead of 489 years, before segmentation, and 44 periods, or bins, in the final histograms.

We found 45,759 MWEs (out of 65,450 entries in the MWE index) in the arcs, or trigram files of the dataset (see research methods, above, for details). 41,366 MWEs of them had a statistically significant trend – an increase or decrease in counts – over the years (Kendall’s $\tau|z| > 3$ or Daniels Test for trend, where Spearman’s $|\rho| > 0.392$; $\alpha = .01$).

The most frequently used expressions were *of which* and *in case* (5% frequency, or 50,000/Million words, over a total of 30 periods – 210 years), while the least frequently used expressions were *bunker buster* and *qassam brigades* (0.122/Million words, over a total of 28 years). Figure 1 plots the normalized frequency versus rank of each expression that was found, and shows that Zipf’s law (Estoup, 1916, as mentioned in Manning & Schutze, 1999), which states that there is a constant relationship between word frequencies and their rank, fits most of the expressions we have explored.

93% of expressions had a sparse histogram, meaning they were used during a rather short period in the dataset (i.e. 90% sparsity corresponds

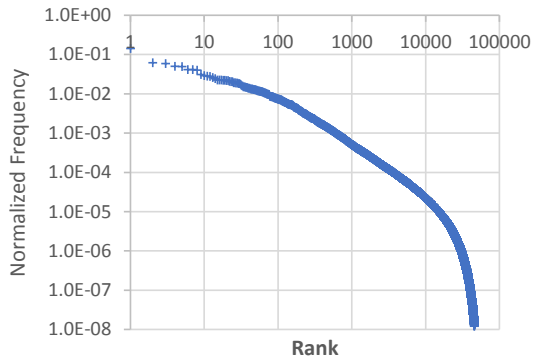


Figure 1: Rank versus Normalized frequency, using logarithmic scales.

to usage during 4 periods – 28 years). These MWEs were mostly named entities, as *Georgia O’Keef*, though some of them were rarely used MWEs (e.g., *Sheath pile*), or new expressions such as *web log*. In order to overcome these problems, we selected only MWEs with a trend that were used for at least 30% of 7-year periods. That step left us with 15,895 MWEs (907 of them with negative trends) that were frequently used across most periods of the dataset, so we could clearly see change in their usage and focus on prevalent expressions. Table 1 shows the 30 expressions with the most increasing usage trends, and Table 2 shows the 30 expressions with the most decreasing usage trends that were found in the dataset.

4.3 Finding Candidate MWEs

In addition to ready-made MWEs found in the dataset, collocations that were not included in the ready-made MWEs list [24] were considered candidate expressions if they passed two thresholds. We set the normalized frequency threshold to $1.22E-08$, which equals the normalized frequency of the least mentioned MWE that was found in the MWE list (Kulkarni and Finlayson, 2011). This threshold represents 0.122 mentions per million words, or 1,359 mentions across the 111 Billion words in the Google Syntactic n-gram dataset (between the years 1701–2008). We also set the sparsity threshold to 4 periods – the shortest period an MWE spans, which equals to 28 years. In order to find only newer expressions, we looked for candidate expressions that started to appear since 1904.

Using these thresholds, we found 4,153 candidate expressions. 2,881 of them had a statistically significant trend ($\alpha = .01$), of which, only 13

showed a decreasing trend. 24 (80%) of the 30 candidate expressions with the most increasing usage trend have MWE characteristics, though some of them are actually professional terms used only in a very specific domain, such as *acoustic energy*, *learning environment*, and *control subject*; However, seven of the candidate expressions were not found in dictionaries⁴, while showing characteristics of a multi-word expression as *Diary entry*, *older adult*, *entry into force*, *emergency entrance*, etc. This may suggest that the two thresholds can be used to find candidate multiword expressions in a multi-year corpus of collocations, as a complement to other existing methods for finding MWEs.

We have also evaluated two other methods that select candidate expressions by taking into account (1) only the normalized frequency values, or (2) only the sparsity values, without taking into account the trend value. We compared the three methods using *precision@k* measure, which allows to track the precision over a range of candidate expressions (collocations) list sizes. As Figure 2 shows, it seems that the best method is to select candidate expressions by sparsity alone while leaving-out proper name expressions.

4.4 Trend Analysis

Before looking at expressions with trends, we looked how expressions with no statistically significant trend behave. We chose expressions that have nearly constant mean number of mentions, and their Kendall’s τ test statistic and Spearman’s ρ are relatively far from statistically significant values.

Two expressions (*collect call* and *lift up*) had no trend and behaved almost as straight lines; other expressions did not behave as a straight horizontal line, as one expects when no trend is reported, however, this fits our expectations from Kendall’s τ and Spearman’s ρ to identify a statistically significant trend only with high confidence ($\alpha = .01$): Expressions with high frequency peak fluctuations (e.g., *white wine*, or *tribes of Israel*) had a trend canceling effect by previous or future fluctuations, in Kendall’s τ equation 2, which is based on the sum of differences. Expressions with a peak in frequency towards the end, as *natural language*, had no trend too since the trend is rather short (the last 48 years of over a period 300 years).

⁴Merriam-Webster dictionary (<http://www.merriam-webster.com/dictionary/>) and Oxford English Dictionary (<http://www.oed.com/>).

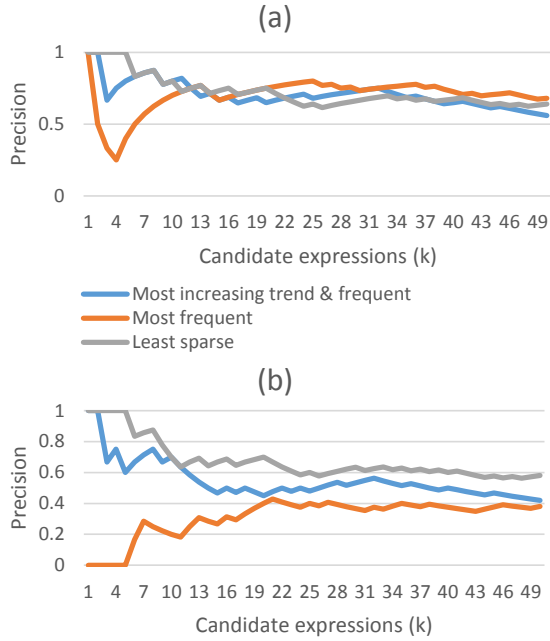


Figure 2: Comparison of the three methods to find candidate expressions, using $Precision@k$ measure. In (a), precision was calculated for all candidate expressions. In (b), precision was calculated after leaving-out proper name expressions (marking them as non-valid expressions).

These results have confirmed the robustness of our tests.

It is noteworthy that some expressions with the most decreasing trends in Table 2 are related to religion (e.g., *revealed religion*, *god almighty*, *Church of Rome*, *St. Peter*, and *high church*). Though our work does not explain language changes, this may be an interesting finding for sociolinguistic researchers, which may indicate a secularization process.

4.5 Top Increasing trends

In order to find old–new relationships between the time-series of some synonymous expressions, we chose the top 30 expressions with the most increasing usage trend, and looked for their historical synonymous expressions in a thesaurus. By visualizing the trends of the synonymous expressions, we could find evidence that later expressions replaced earlier ones. We found synonymous expressions in a thesaurus for 8 out of the 30 expressions in Table 1: *in practice*, *better off*, *talk about*, *go wrong*, *In fact*, *for instance*, *police officer* and *on and off*. However, we did not find

Increasing trends	Kendall's	Spearman's
	τ	ρ
in turn (r)	9.568	1.000
in practice (r)	9.528	1.000
better off (j)	9.528	1.000
think about (v)	9.507	1.000
work through (v)	9.497	0.999
white woman (n)	9.497	1.000
human being (n)	9.487	0.999
talk about (v)	9.487	0.999
written record (n)	9.447	0.999
united kingdom (n)	9.437	0.999
rule of law (n)	9.406	0.999
take into account (v)	9.406	0.998
two dozen (n)	9.396	0.998
rather than (r)	9.386	0.998
go wrong (v)	9.386	0.998
human activity (n)	9.376	0.998
in fact (r)	9.366	0.997
Cambridge university (n)	9.366	0.999
bring together (v)	9.346	0.997
san Antonio (n)	9.335	0.998
critical analysis (n)	9.335	0.998
for instance (r)	9.325	0.995
end on (r)	9.325	0.997
life form (n)	9.325	0.997
police officer (n)	9.325	0.997
medical history (n)	9.315	0.998
run by (v)	9.305	0.997
conflict of interest (n)	9.305	0.998
per year (r)	9.295	0.996
on and off (r)	9.295	0.997

Table 1: 30 expressions with the highest increasing usage trend. (n – noun phrase; v – verb phrase; j – adjective; r – adverb; o – other).

synonymous expressions in our ready-made MWE list for *in practice* and *better off*. None of the synonymous expressions for the remaining 6 expressions had a statistically significant decreasing usage trend. Here are some detailed examples:

Talk about is “[...] often used colloq. to contrast something already mentioned with something still more striking; [...]” (Talk, v., 2015). Its synonym expressions are *talk of*, as well as *speak of* – a synonym not mentioned in Oxford English Dictionary. Figure 3 shows that *speak of* is more widely used than *talk about* since it may have additional meanings, as stating another example to the discussion, where *talk about* and *talk of* are used only

Decreasing trends	Kendall's	Spearman's
	τ	ρ
take notice (v)	-9.184	-0.994
no more (r)	-9.164	-0.991
as much (o)	-9.143	-0.993
king James (n)	-9.103	-0.989
ill nature (n)	-9.062	-0.990
according as (j)	-9.062	-0.988
root out (v)	-8.941	-0.985
think piece (n)	-8.799	-0.987
high church (n)	-8.718	-0.979
of it (r)	-8.718	-0.976
make happy (v)	-8.658	-0.979
fourth part (n)	-8.658	-0.965
St. peter (n)	-8.638	-0.979
church of rome (n)	-8.597	-0.973
ought to (v)	-8.557	-0.972
good nature (n)	-8.557	-0.971
god almighty (n)	-8.536	-0.975
give ear (v)	-8.476	-0.974
law of nature (n)	-8.476	-0.948
let fly (v)	-8.415	-0.973
bring forth (v)	-8.415	-0.968
build upon (v)	-8.354	-0.969
perpetual motion (n)	-8.334	-0.971
revealed religion (n)	-8.334	-0.940
many a (j)	-8.314	-0.968
states general (n)	-8.314	-0.966
take care (v)	-8.294	-0.951
as many [as] (j)	-8.273	-0.956
take pains (v)	-8.273	-0.940
nemine contradicente (r)	-8.253	-0.957

Table 2: 30 expressions with the most decreasing usage trend. (*n* – noun phrase; *v* – verb phrase; *j* – adjective; *r* – adverb; *o* – other).

to contradict a point in the discussion. Though *talk of* has no significant decreasing trend, it shows a decline along the 20th century.

The expression [*to*] *go wrong* has several meanings: It could mean to take a wrong way, either literally, in mistake, or morally. It could also mean that an event “[...] can happen amiss or unfortunately[, when something broke-down, or when food [...] get[s] into bad or unsound condition [...]” (Wrong, adj. and adv., 2015). It has many synonym expressions; in Figure 4 we compare it with synonyms we found in the ready-made MWE list (Kulkarni and Finlayson, 2011): *break down* (1837), *go bad* (1799) and *go off* (1695).

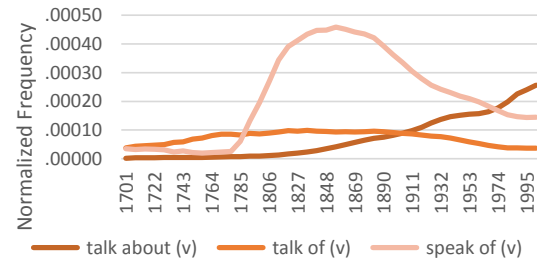


Figure 3: Comparison between *talk about*, *talk of* and *speak of*.

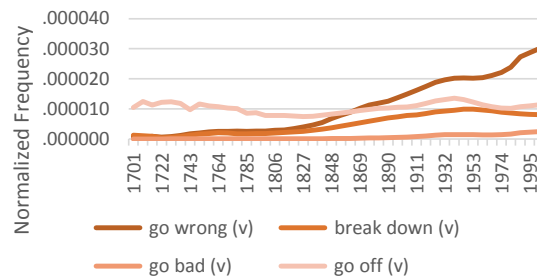


Figure 4: Comparison of *go wrong* with its synonymous expressions.

In fact (dated 1592) is defined as “in reality, actually, as a matter of fact. Now often used parenthetically as an additional explanation or to correct a falsehood or misunderstanding (cf. in point of fact at Phrases 3)” (Fact, n., int., and adv. [P2], 2015). In Figure 5 we compare it with synonyms we found in the ready-made MWE list (Kulkarni and Finlayson, 2011): *smatter of fact* (1922), *in effect*, *in truth* (1548), *in esse[n]ce*], and *de facto* (Really or actually [adverb], 2015).

The expression *on and off* has an earlier synonym expression: *off and on* (On and off, adv., adj., and n., 2015), as shown in Figure 6. Both expressions have statistically significant increase trends, while *on and off* exceeds *off and on* since around 1921.

4.6 Top Decreasing trends

Similar to the previous section 4.5, we chose the top 30 expressions with the most decreasing usage trend, and looked for their historical synonymous expressions in a thesaurus. Again, we saw evidence that later expressions replace earlier ones.

In total, we found synonymous expressions in a thesaurus for 7 out of the 30 expressions in Ta-

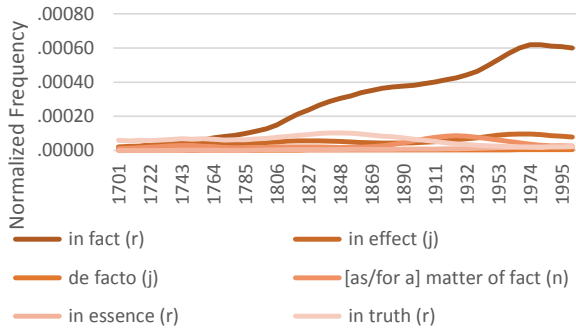


Figure 5: Comparison of *in fact* with its synonymous expressions.

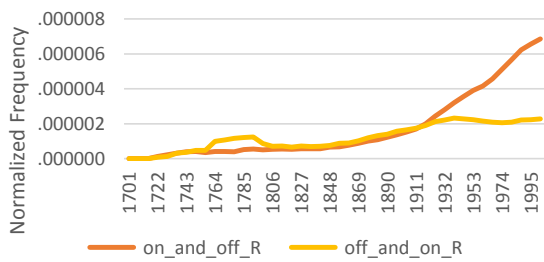


Figure 6: Comparison between *on and off* and *off and on*.

ble 2: *Let fly, take notice, give ear, law of nature, good nature, ought to* and *no more*. However, we did not find synonymous expressions for *good nature* in our ready-made MWE list, to compare with. All of the synonymous expressions for the remaining 6 expressions had a statistically significant increasing usage trend, hinting that old–new relationships exist between them. In addition, expressions with decreasing trends were often found in Oxford Online Dictionary⁵ as an obsolete, rare, or poetic expressions. Here are two examples:

The expressions *take notice* and *give ear* could also be phrased as *pay attention* or *take heed* (Notice, n., 2015). The expression *pay attention* has an increasing trend, and may partially explain the decrease of *take notice*, as shown in Figure 7. The drastic decrease in usage of the expression *take notice* could also be explained by single-word synonyms as *note*, *notice*, and *listen*, which we did not compare to.

Though *no more* has several meanings, we found in the MWE list (Kulkarni and Finlayson,

⁵<http://www.oed.com/>

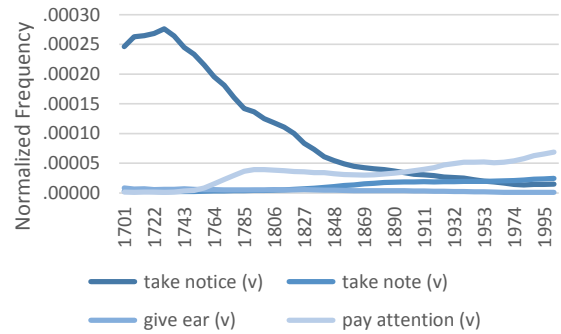


Figure 7: Comparison of expressions *take notice*, *take note*, *give ear* and *pay attention*.

2011) only synonyms in the sense of never again or nevermore: *never again* and *no longer* (Never again, 2015):

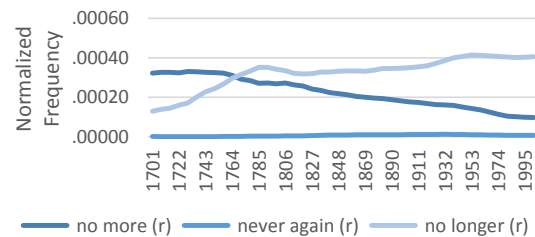


Figure 8: Comparison of *no more*, *no longer*, and *never again*.

5 Discussion & Conclusions

We explored the change in Multiword expressions (MWEs) usage, or functionality over the years. By visualizing the trends of synonymous expressions, we found evidence to our hypothesis that old–new relationship exists between some expressions: We found synonymous expressions with an increasing usage trend for all 6 expressions with decreasing usage trends, though we did not find decreasing usage trends for synonymous expressions of expressions with increasing usage trends. We found that some expressions with the most decreasing trends are related to religion, which might interest sociolinguists.

We showed that it is possible to find new MWEs in a historical collocations corpus using either normalized frequency or sparsity thresholds, as seven of the 24 candidate expressions were found to be

metaphoric phrases not included in dictionaries⁶. Using normalized frequency was better, on average, as a criterion to find any type of candidate expressions, whereas using sparsity was better if one is not interested in proper name expressions. Expressions in the MWE list (Kulkarni and Finlayson, 2011) were mentioned in the Google Syntactic-Ngrams dataset for at least 28 years in a row. This may suggest a minimum period lexicographers can test an expression against before entering it into a dictionary or thesaurus.

In the future, it is possible to tweak Kendall's τ coefficient, especially equation 2, so a short-term trend towards the end of the time-series would also be recognized as statistically significant. Future work may also improve the methods for finding MWEs by introducing flexibility in the expression structure, and by using synonym words replacement. These would assist lexicographers to track the evolution of human language. A usage trend may also be used as a feature by an MWE extraction algorithm; the historical perspective of an expression usage may be valuable for identifying stable expressions, while filtering out short-term collocations.

References

- U.S. Environmental Protection Agency. 1974. Guideline for the evaluation of air quality trends: Guideline series (oaqps no. 1.2-014). Technical report, U.S. environmental Protection Agency, Office of Air Quality Planning and Standards, Monitoring and Data Analysis Division. Available from: <http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=9100FOJ5.txt>.
- H. Al-Haj and Shuli Wintner. 2010. Identifying multi-word expressions by leveraging morphological and syntactic idiosyncrasy. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, pages 10–18, Beijing, Aug 2010.
- B. T. S. Atkins and M. Rundell. 2008. *The oxford guide to practical lexicography*. Oxford University Press, Oxford.
- T. Baldwin and S. N. Kim, 2010. *Handbook of Natural Language Processing*, chapter Multiword Expressions, pages 267–292. CRC Press, Boca Raton, USA, 2nd edition. N. Indurkha, F. J. Damerau, editors; DOI: 10.1038/nbt1267.
- A. Buerki. 2013. Automatically identifying instances of change in diachronic corpus data. Presentation at the Corpus Linguistics 2013 conference, Jul 22-26.
- Nicoletta Calzolari, Charles J. Fillmore, Ralph Grishman, Nancy Ide, Alessandro Lenci, Catherine MacLeod, and Antonio Zampolli. 2002. Towards best practice for multiword expressions in computational lexicons. In *Proceedings of the third LREC (LREC 2002)*, pages 1934–1940, Las Palmas, Canary Islands, Spain.
- H. E. Daniels. 1950. Rank correlation and population models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(2):171–191.
- J. B. Estoup. 1916. *Gammes Stenographiques*. Institut Stenographique de France, Paris, 4th edition.
- Fact, n., int., and adv. [P2]. 2015. In *Oxford English Dictionary*. Oxford University Press. [cited 2015 Mar 10]. Available from: <http://www.oed.com/view/Entry/67478>.
- A. Fazly, P. Cook, and S. Stevenson. 2009. Unsupervised type and token identification of idiomatic expressions. *Computational Linguistics*, 35(1):61–103.
- Dylan Glynn. 2010. Synonymy, lexical fields, and grammatical constructions: Developing usage-based methodology for cognitive semantics. In H. J. Schmid and S. Handl, editors, *Cognitive Foundations of Linguistic Usage Patterns*, pages 89–118. Mouton de Gruyter, Berlin.
- Yoav Goldberg and Jon Orwant. 2013. A dataset of syntactic-ngrams over time from a very large corpus of english books. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics (*SEM 2013)*.
- Google. 2009. Google ngram viewer, total counts file for english one million corpus, version 20090715. Available from: <http://storage.googleapis.com/books/ngrams/books/googlebooks-eng-1M-totalcounts-20090715.txt> [Accessed: Aug 27, 2014].
- K. L. Gray. 2007. *Comparison of Trend Detection Methods*. Ph.D. thesis, University of Montana, Department of Mathematical Sciences, Missoula, MT. 86 pages.
- B. Juska-Bacher and C. Mahlow. 2012. Phraseological change – a book with seven seals? tracing back diachronic development of german proverbs and idioms. In M. Durrell, S. Scheible, and R. J. Whitt, editors, *Volume of Corpus linguistics and Interdisciplinary perspectives on language*. Gunter Narr, Tbingen, Germany.
- M. Kalla. 2006. A diachronic semiotic analysis of words signifying 'friendship' in hebrew. Master's thesis, Ben-Gurion Univ. of the Negev, Dept. of Foreign Languages and Literatures, Beer-Sheva, Israel. 96 pages.

⁶Merriam-Webster dictionary (<http://www.merriam-webster.com/dictionary/>) and Oxford English Dictionary (<http://www.oed.com/>).

- Tom Kenter, Pim Huijnen, Melvin Wevers, and Maarten de Rijke. 2015. Ad hoc monitoring of vocabulary shifts over time. In *CIKM 2015: 24th ACM Conference on Information and Knowledge Management*. ACM, October.
- N. Kulkarni and M. A. Finlayson. 2011. jmwe: A java toolkit for detecting multi-word expressions. In *Proceedings of the 2011 Workshop on Multiword Expressions (ACL 2011)*, pages 122–124, Portland, OR.
- C. Liebeskin, Ido Dagan, and J. Schler. 2013. Semi-automatic construction of cross-period thesaurus. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2013)*, pages 29–35, Sofia, Bulgaria, Aug 8. The Association for Computational Linguistics.
- C. Mahlow and B. Juska-Bacher. 2011. Exploring new high german texts for evidence of phrasemes. *Journal for Language Technology and Computational Linguistics*, 26(2):117–128.
- C. D. Manning and H. Schütze, 1999. *Foundations of Statistical Natural Language Processing*, chapter Introduction, pages 23–29. The MIT Press, Cambridge, Massachusetts; London, England.
- C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.
- Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182. DOI: 10.1126/science.1199644.
- Never again. 2015. In *Oxford English Dictionary*. Oxford University Press. [cited 2015 Feb]. Available from: <http://www.oed.com/view/th/class/96637>.
- Notice, n. 2015. In *Oxford English Dictionary*. Oxford University Press. [cited 2015 Mar 10]. Available from: <http://www.oed.com/view/Entry/128591#eid933873046>.
- On and off, adv., adj., and n. 2015. In *Oxford English Dictionary*. Oxford University Press. [cited 2015 Mar 10]. Available from: <http://www.oed.com/view/Entry/131310>.
- C. Ramisch. 2013. *A Generic open framework for multiword expressions treatment: from acquisition to applications*. Ph.D. thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil & Grenoble, France.
- Really or actually [adverb]. 2015. In *Oxford English Dictionary*. Oxford University Press. [cited 2015 Mar 10]. Available from: <http://www.oed.com/view/th/class/82683>.
- Talk, v. 2015. In *Oxford English Dictionary*. Oxford University Press. [cited 2015 Mar 10]. Available from: <http://www.oed.com/view/Entry/197246>.
- Yulia Tsvetkov and Shuli Wintner. 2011. Identification of multi-word expressions by combining multiple linguistic info sources. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pages 836–845, Edinburgh, Scotland, UK, Jul 27–31. Association for Computational Linguistics.
- Wrong, adj. and adv. 2015. In *Oxford English Dictionary*. Oxford University Press. [cited 2015 Mar 10]. Available from: <http://www.oed.com/view/Entry/230802>.