Probabilistic Measures for Diffusion of Linguistic Innovation: As Seen in the Usage of Verbal "Nok" in Thai Twitter

Nozomi Yamada. Pittayawat Pittayaporn.
254 Phayathai Road, Pathumwan, Bangkok 10330 Thailand
Southeast Asian Linguistics Research Unit and
Department of Linguistics, Faculty of Arts, Chulalongkorn University
y.nozomi320@gmail.com
Pittayawat.P@chula.ac.th

Abstract

The existence of several SNS (social networking service) such as Twitter accelerates the diffusion process of language change. In this paper, we examine the diffusion of the innovative verbal usage of *nók* in Thai Twitter. We collected more than 25 millions tweets and adopted not only word frequency but also three probabilistic measures of analysis: conditional probability, PMI and cosine similarity of word embeddings. The result of these three probabilistic measures show the stability of the innovation regardless of decrease of word frequency. These facts support the idea that the innovation *nók* is lexically established in Thai language. Most importantly, it shows that the three probabilistic measures can be used to quantify diffusion of linguistic innovation regardless of its polysemy.

1 Introduction

Twitter is one of the most popular social networking services in Thailand. Though there is no official demographic profile, some online statistics websites like *we are social*[1] rank Twitter as the 3rd most popular SNS in Thailand behind Facebook and Instagram. Not only is it a large, free sources of relatively casual language used in daily communication (Crystal, 2006), but also potential space for examining early stages of language change. As networks in Twitter mainly consist of weak ties characterized by occasional contacts and lack of emotional bonding (Virk, 2011), linguistic innovations can spread quickly in Twitter, making it possible to observe complete propagation of language change in a short period of time.

An interesting and methodologically challenging case study is that of the verbal *nók* in Thai, an innovation gaining currency among Thai speakers. Originally a noun that means "bird", at present, the word is also used as a slang meaning "to fail to achieve one's expectation", especially used in the context of love or flirting. Although it is not clear when *nók* first came to be used as a verb, it was already popular to some extent among transgender women and gay men in 2014, and was commonly used among TV personalities by 2015.

This paper explores how Twitter data can be used to analyze the diffusion of an innovative lexical usage by taking the example of verbal *nók*. This innovation is chosen because it is a case of polysemy. Unlike cases in which a new variant propagates at the expense of an old one (Nevalainen and Raumolin-Brunberg, 2016), the verbal *nók* is not in competition with any other word. More specifically, the polysemy poses two challenges: how to detect and separate the innovative usage from the original usage for data processing, and how to quantify its progress through the linguistic system.

Therefore, this paper shows that changes in conditional probability, PMI, and cosine similarity of word embeddings are better measures for diffusion progress than word frequency. These measures also show that the verbal *nók* has been established as a new usage and and is broadening its meaning in Thai language.

[1] https://www.slideshare.net/DataReportal/digital-2019-thailand-january-2019-v01

33rd Pacific Asia Conference on Language, Information and Computation (PACLIC 33), pages 271-278, Hakodate, Japan, September 13-15, 2019

2 Literature Review

2.1 Linguistic Innovation and Diffusion

Linguistic innovations are the ultimate origins of language changes (Milroy and Milroy, 1985), and such innovative usage of *nók* can be viewed as a case of lexical innovation. As Sornig (1981) explains, there are many kinds of lexical innovations: new word for new concept, new competitive word for existing word, new meaning for existing word, and so on. The case of our study verbal *nók* is an example of "new meaning for existing word".

Following Rogers (1962) who demonstrates that the diffusion of innovations is often represented as S-curve (logistic curve / sigmoid curve), Trudgill (1974) and Milroy and Milroy (1985) extended the theory to linguistic innovations, showing that they diffuse in the same manner as other social phenomena. More recently, Yang (2000) make probabilistic models for language change to explain how competing variations of word order in Old English and Old French were diffused and decayed. Blythe and William (2012) and Kauhanen (2017) put an emphasis on genetic replicating process in utterances, and make several mathematical models for explaining language change and S-curve. Ishii, et al. (2012) focus not only internal networks but also the relationship between external effects and utterance in order to give an account for short-term phenomenon as well as long-term propagation.

While these studies mainly focus the mechanism of diffusion, namely "how innovation is propagated", there are other studies that focus "whether innovation is diffused or not". As previous studies such as Nation and Waring (1997) and Piantadosi (2014) shows, word frequency is one of the most inteligible criteria for measuring generality of words. Metcalf (2004) and Barnhart (2007) thus introduce a scale for measuring acceptance of a new word by using word frequency. Phillips (2006) asserts that the diffusion of lexical innovation also becomes S-curve in the same way as other linguistic innovations by using word frequency. Hilpert and Gries (2008), using historical corpora, claim that diachronic change of word frequency directly indicates that the language change is in progress.

2.2 Language Change and Twitter

Social media like Twitter is an exciting domain for investigating the propagation of language change. Viewing language change as a result of diffusion of a speaker innovation (Milroy and Milroy, 1985), social networking services allow us to observe the rates at which linguistic innovation diffuses through speech communities and through linguistic systems. For example, Maybaum (2013) investigated several new words related to Twitter such as *tweeps, tweeple* by using big size of tweet data and showed that there is a tendency to be S-curve. However, this case is a type of "new word for new concept", which is not the case of *nók*.

Kershaw, Rowe and Stacey (2016) investigated innovation acceptance in twitter by using measuring scale of Metcalf (2004) and Barnhart (2007) and showed significance of word frequency. On the other hand, Yamanouchi and Komatsu (2014) focus not only word frequency, but also stochastic process of utterance and alpha-stable distribution of probability of each word. While word frequency may easily fluctuate under the influence of external events or randomness, the indices that determines distribution of probability are more stable. They proved the fact by sampling data from Twitter.

3 Data and pre-processing

The data we used is tweets written in Thai language from January 2012 to December 2018 (Table 1). First, we collected about 1000 - 2500 tweets per day (data set A) containing the word *nók*. This data is used for two analyses: conditional probability and word embeddings. Next, we collected about 10000 - 25000 random tweets per day (data set B) for three analyses: word frequency, PMI and word embeddings. In order to prevent from being biased, we collected tweets every 10 minutes.

The most crucial step in data pre-processing is to distinguish cases of the innovative verbal *nók* from cases of the original nominal usage. The most reliable heuristic is its co-occurrence with negator *mâi* or auxiliary verb. As the verbal *nók* can only have the innovative meaning 'to fail achieve one's expectations', the two occurrence patterns also distinguish between the old and new meanings. Table 2 gives examples of the most common syntactic structures

Year	A: Tweets with *nók*	B: Random Tweets
2012	118,799	0
2013	476,365	2,529,665
2014	425,421	3,732,020
2015	395,334	3,153,596
2016	778,243	3,434,185
2017	1,070,668	5,152,559
2018	891,636	3,556,596
total	4,156,466	21,558,621

Table 1: The number of collected tweets

in Thai.

Structure	Example	Gloss
S V	*phŏm pai*	I go
S Adj	*phŏm hĭw*	I hungry
S NEG V	*phŏm mâi pai*	I not go
S NEG Adj	*phŏm mâi hĭw*	I not hungry
S AUX V	*phŏm cà pai*	I will go
S Cop N	*phŏm pen nók*	I be bird
S NEG Cop N	*phŏm mâi châi nók*	I not be bird

Table 2: Most common syntactic structures in Thai

The copular verbs *pen* and *châi* are needed when the sentence is nominal sentence as shown in the last two examples. In other words, the collocations *mâi nók* or *AUX nók* will never occur as long as *nók* is occurs as nominal meaning "bird". Therefore, to make a list of co-occurrences proved essential.

In order to find these co-occurrences, we tokenized all tweets with the python toolkit `PyThaiNLP 2.0.3`, and the Maximum-Matching (MM) algorithm. MM algorithm requires a vocabulary set (dictionary) for tokenization, and we can control it. Since we wanted to locate words preceding/following *nók*, we removed all compound words containing *nók* such as *nókphirâap* ("pigeon"), from the vocabulary set beforehand.[2]

In addition, there are some tweets that include repetition of the same characters or words in order to exaggerate, such as "aaaaraaaaiiii", "nóknóknóknóknók". Since these repetitions make it more difficult for a program to tokenize, we detected them by using regular expressions, then condensed them before tokenization.

[2]then, *nókphirâap* is tokenized as *nók* and *phirâap*

4 Measures of Diffusion

4.1 Word Frequency

As mentioned above, word frequency is one of the most popular methods for measuring diffusion of innovation. We calculated word frequency (per 10000 words) by counting tokens of *nók* as well as all tokens for each month, then traced the diachronic change. However, the word *nók* is not a new word but a polyseme, so word frequency alone does not indicate how often the innovative verbal *nók* is used. We thus needed other methods to separate the two usages and normalize them.

4.2 Conditional Probability of Bigrams

Bigrams are an important methods in computational linguistics especially in building language models. For our purposes, bigrams are employed in detecting the syntactic structure of the sentence is and how often that structure occurred. We mentioned in section 3 that verbal sentences and nominal sentences take different structures. As such, identifying collocations can help to distinguish the two meanings. However, calculation of collocation must be normalized so that we could compare diachronically regardless of the size of the data. We thus defined conditional probability for preceding word $P_{pre}(w_i|nok)$ and conditional probability for following word $P_{fol}(w_i|nok)$ as follows:

$$P_{pre}(w_i|nok) = \frac{C(w_i, nok)}{\sum_w C(w, nok)} \quad (1)$$

$$P_{fol}(w_i|nok) = \frac{C(nok, w_i)}{\sum_w C(nok, w)} \quad (2)$$

where $C(w_i, nok)$ is the total number of co-occurrences of a word w_i and *nók* in all tokens[3], $\sum_w C(w, nok)$ is the total number of co-occurrences containing *nók* as the second word of the bigram. For example, the conditional probabilities for words A, B, C in the sentence "*nók A B nók B C nók A*" are given by following calculations:

$$\begin{array}{ll} P_{pre}(A|nok) = 0 & P_{fol}(A|nok) = 2/3 \\ P_{pre}(B|nok) = 1/2 & P_{fol}(B|nok) = 1/3 \\ P_{pre}(C|nok) = 1/2 & P_{fol}(C|nok) = 0 \end{array}$$

[3]not including white space and punctuation

4.3 Tweet-Level PMI

Another probabilistic measure is Pointwise Mutual Information (PMI) at tweet level. PMI is a measure of the independency of two words (Jurafsky and Martin, 2014), the degree of how often (or not) the two words co-occur. PMI is similar to conditional probability of bigrams, though the method of normalization differs slightly. Moreover, with conditional probability, we did not consider which tweet the bigram comes from. In other words, we dealt with all tweets as one text. Here, we define tweet-level PMI of bigrams as:

$$\mathrm{PMI}(w_i, w_j) = \log_2 \frac{p(w_i, w_j)}{p(w_i)p(w_j)} \qquad (3)$$

where $p(w_i, w_j)$ is the probability that one tweet contains a co-occurrence of the bigram (w_i, w_j), in short, the proportion of the tweets that contain the target bigram. $p(w)$ is the probability of occurrence of word w in any given tweet. The normalization factor (given in the denominator) is the possibility of occurence for each word. In this case, either of the two words is *nók*. Since every tweet in data set A contains the word *nók* and therefore, $p(nok) = 1$, we used only data set B (random tweets) for this PMI calculation.

4.4 Cosine Similarity of Word Embeddings

The third probabilistic measure is cosine similarity of word embeddings. Though a word embedding itself does not provide a direct probability, the methods of obtaining word embeddings, such as SVD or word2vec, are based on distribution of words. We thus refer to it as "probabilistic" in a broad sense. Several studies such as Hamilton, Leskovec and Jurafsky (2016), Bamler and Mandt (2017), Baitong, Ying and Feicheng (2018) reveal that comparison of word embeddings derived from various periods of historical corpora can, in fact, reveal language change. Moreover, their studies also show that a word embedding of a polyseme is located between the embeddings of its two meanings. In other words, we can observe the propagation of language change by measuring whether the cosine similarity between the word *nók* and another word that means "to fail to achieve one's expectation" is rising or not.

We employed the word2vec toolkit `gensim 3.7.1` in computing 300-dimension word embeddings for each month from 2014 to 2018. In order to obtain not only the word embedding of *nók* but also the word embeddings of various words for comparison, we first combined two data sets: data set A (tweets that contain *nók*) and data set B (random tweets). For all word embeddings, we used a CBOW algorithm with symmetric windows of size 5, and iterated for 3 epochs. Though skip-gram algorithms are more popular at present, we used a CBOW algorithm as it works better with frequently occurring words (Naili, Chaibi, and Ghezala, 2017).[4]

5 Results

5.1 Word Frequency

Before addressing the frequency of *nók*, we first performed a preliminary test of a selected set of basic words -frequencies of basic words should be stable diachronically- to check that our data set was sufficiently sized and void of bias. We chose three frequent words: *mâi* "not", *pen* "be" (copula) and *tham* "do". Figure 1 gives a plot of frequencies per 10,000 tokens of each word in data set B (random tweets) from 5/2013 to 10/2018

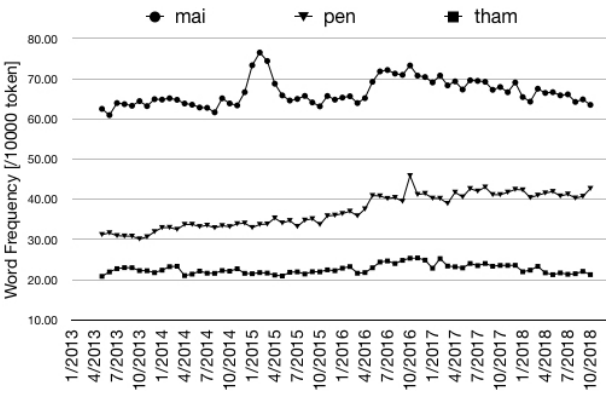

Figure 1: Word frequency of 3 words in data set B

The frequency of *mâi* fluctuates within the period examined, while *pen* is increases gradually, and *tham* is almost stable. Though these three words display slightly different patterns, none of them show an abrupt change. We can thusly conclude that our set is not biased.

[4] *bird* is 20th on the Swadesh list

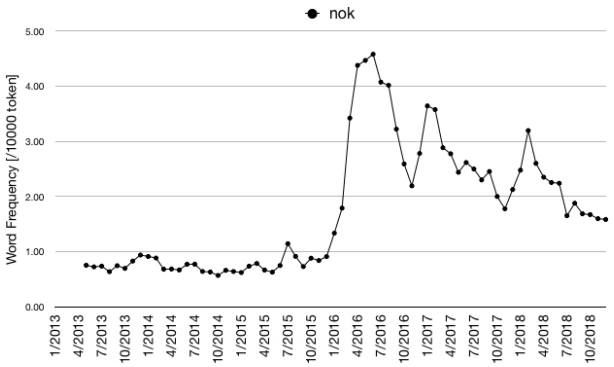

Figure 2: Word frequency of *nók* in data set B

Next, let us compare the frequency of *nók* shown in Figure 2 which, contrastively to the frequencies of the basic words, increases abruptly at the beginning of 2016 and reaches its peak in the middle of the same year at more than 4 times its original value. After that, it resembles exponential decay, falling to only 1.6 times the original value. From this result, the innovation seems to be disappearing rather than diffusing. Though this abrupt change may be evidence of linguistic innovation, we cannot make this determination based only on this data, as this word frequency is the sum of both original *nók* and verbal *nók*. Since our data is too large to check one by one, we cannot obtain how much the word frequency of verbal *nók* is.

5.2 Probabilistic Measures

The three probabilistic measures to be discussed here demonstrate a contrastive aspect to that of word frequency. First is the calculation of P_{pre} and P_{fol}, for which we selected three words capable of distinctively indicating syntactic structure. Tables 3 and 4 lists these words.

Word	POS	Meaning	Note
mâi	Adv	"not"	negation
cà	AUX	"will"	V follows
jàa	AUX	"Don't"	imperative

Table 3: 3 words selected for P_{pre} calculation

The three words for P_{fol} do not directly indicate the status of *nók* as a verb as much as the words for P_{pre}; however, they tend to follow verbs in Thai. Figure 3 and Figure 4 are transitive graphs of conditional probability P_{pre} and P_{fol}.

Word	POS	Meaning
laéw	Adv	"already"
ìik	Adv	"again", "more"
talòt	Adv	"always"

Table 4: 3 words selected for P_{fol} calculation

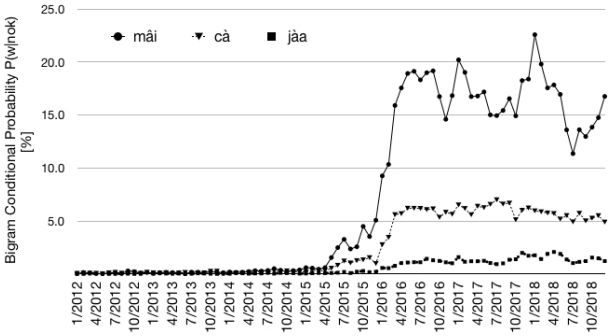

Figure 3: Conditional probability for the preceding word

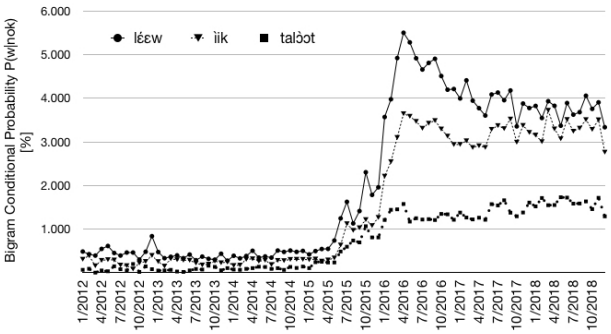

Figure 4: Conditional probability for the following word

Figure 3, of course, demonstrates an increase in each co-occurrence after 2015. Conditional probability forms an S-curve, hardly decaying after reaching its peak.

Figure 4 displays similar patterns. Though only *laéw* decays after reaching its peak, its value later becomes stable in the same way as the other two words, *ìik* and *talòt*.

A similar shape occurs, as well, with calculation of tweet-level PMI for the same two words.

Figure 5 shows PMI(*mâi*, *nók*) and PMI(*cà*, *nók*). Notably, some points are not plotted as the log of zero will equal negative infinity. Both curves

Word	POS	Meaning
mâi	Adv	"not"
cà	AUX	"will"

Table 5: 2 words selected for PMI calculation

abruptly increase in 2016 as was the case with conditional probability. Though they seem to decrease gradually, neither mirrors the decreasing pattern of word frequency and can be considered more stable.

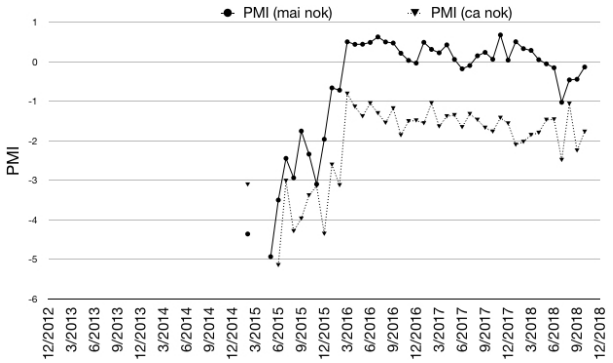

Figure 5: PMI(*mâi*, *nók*) and PMI(*cà*, *nók*)

The third measure is cosine similarity of word embeddings. We selected the two words below and measured cosine similarity between each of these and *nók*. As shown in Table 2, in Thai, verbs and adjectives will appear in the same structures. Thus, we can regard these two words as synonyms of verbal *nók*.

Word	POS	Meaning
plâat	Verb	"to miss", "to fail"
sĭacai	Adj	"(to feel) sorry"

Table 6: 2 words selected for cosine similarity calculation

Figure 6 and 7 show diachronic change of the cosine similarities for each pair. Similarity before 2015 is near zero, then after 2015, it forms an S-curve and the value hardly decays, even after reaching its peak. This means that the new meaning of the verbal *nók* still remains.

6 Discussion

There are obvious differences between measures of word frequency and the three probabilistic measures. We can surely conclude that people have been using *nók* less frequently following its boom in 2016; however, this does not simply mean that the lexical innovation is disappearing. According to Figure 2, word frequency in later 2018 is less than half of that of early 2016. only the word frequency of the new meaning of *nók* decreases while frequency of the old usage remains constant, conditional probability and PMI must decrease as well. On the contrary, the result indicates constancy of both, meaning the proportion of verbal *nók* to total use of *nók* on Twitter is unchanged, regardless of decreases in word frequency itself. Comparing conditional probability and PMI, conditional probability is more convenient for use as its measure requires only tweets containing *nók*, while, for PMI, random tweets must be gathered for calculating the probability $p(nok)$.

This constancy is true of cosine similarity as well, and since cosine similarity can indicate the meaning of the word more directly than conditional probability or PMI, the result is more convincing. The co-

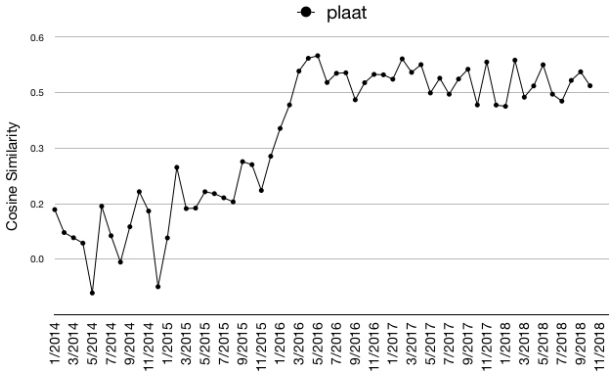

Figure 6: Cosine similarity between *nók* and *plâat*

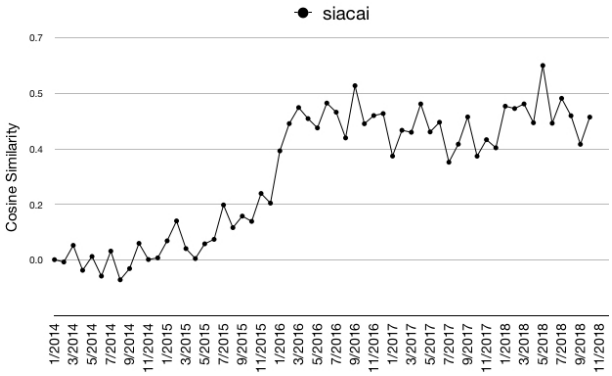

Figure 7: Cosine similarity between *nók* and *sĭacai*

sine similarity is rising throughout 2015, and finally, it becomes stable with no decrease, indicating continued use *nók* in the same context as the two compared words based on the fact that, if the usage is not already established within the linguistic system, cosine similarity would be expected to decay back to zero. According to these results, we can conclude that the lexical innovation *nók* has been established in the linguistic system of Thai.

However, our study has several limitations. First, though our results have revealed characteristics of language change on Twitter, we cannot discern the total acceptance rate. Even if a lexical innovation is already established on Twitter with a constant probability, this does not entail that every Twitter user accepts the innovation. Since this lexical innovation is of type "new meanings for existing words" and not of "new competitive words for existing words", it is impossible to measure 0-100 % acceptance rates from the beginning. We must, therefore, take an "ensemble average" by sampling a nascent language change within the linguistic system and its diffusion pattern.

Second, it is fortunate that *nók* is a polyseme of verb and noun as syntactic structure differs between nominal sentences and verbal sentences, and thusly, conditional probability and PMI containing frequently occurring grammatical words (i.e. negators or auxiliary verbs) can be used. However, if there were a polyseme that is morphosyntactically identical, differentiation may prove to be much more difficult. In that case, we would be forced to use less frequent collocation than that of grammatical words.

As well, we have not analyzed what kinds of mechanisms are present. In other words, we did not show how innovation was propagated, only whether innovation has been diffused or not. There must be at least two factors for the propagation mechanisms: internal networks and external effects. As we reviewed in section 2, there have been many previous studies accounting for diffusion mechanisms through use of various models. In the future, we plan to build from the current research and explore the mechanism of *nók* with such models.

Incidentally, we found another phenomenon occurring in the diffusion of *nók*. The innovative meaning of *nók* in the beginning is just "to fail to flirt" and applies to both men and women. After the innovation had been diffused to the masses, the meaning broadened. In other words, it came to be used as a more generalized verb. Figure 8 gives the PMI of *nók* and *bàt* "ticket". This figure shows that the first appearance and peak for this pair are delayed in comparison to other pairs, and that use is still increasing. Additionally, we found many nouns following *nók*, such as "thing", "live", "giveaway", in the data, suggesting greater favorability toward a wider array of environments and, therefore, a more general meaning. This broadening of meaning also supports the idea that innovative *nók* has been established.

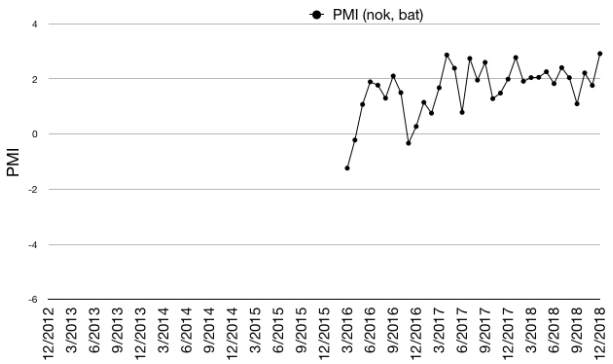

Figure 8: PMI(*nók*, *bàt*)

7 Conclusion

The results indicate that the word frequency of *nók* is now decreasing, while three probabilistic measures are stable over time. This fact supports the idea that the lexical innovation has been established in linguistic system.

The most significant point is that the three measures we adopted can deal with polysemy, unlike word frequency. These measures can be used to quantify diffusion of linguistic innovation regardless of its polymsemy. Though this study examined only one case in Thai, the methods employed here are universally applicable with potential to be extended to other languages as well.

Acknowledgement

We are deeply grateful to Dr. Attapol Thamrongrattanarit for offering good hints for analyses and our colleague Ashley Laughlin for correcting proofs.

References

Baitong Chen, Ying Ding and Feicheng Ma. 2018. Semantic word shifts in a scientific domain. *Scientometrics* 117:211–226

Bamler, R. and Mandt, S. 2017 August. Dynamic word embeddings. In *Proceedings of the 34th International Conference on Machine Learning-Volume* 70 (pp. 380-389). JMLR. org.

Barnhart, D.K. 2007. A calculus for new words. *Dictionaries: Journal of the Dictionary Society of North America*, 28(1), pp.132-138.

Blythe, Richard A., and William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language*, 269-304.

Crystal, D. 2006. *Language and the Internet*. Cambridge, Cambridge University Press.

Hamilton, W.L., Leskovec, J. and Jurafsky, D. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv*:1605.09096.

Hilpert, M. and Gries, S.T. 2008. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing*, 24(4), pp.385-401.

Ishii, A., Arakaki, H., Matsuda, N., Umemura, S., Urushidani, T., Yamagata, N. and Yoshida, N. 2012. The 'hit'phenomenon: a mathematical model of human dynamics interactions as a stochastic process. *New journal of physics*, 14(6), p.063018.

Jurafsky, D. and Martin, J.H. 2014. *Speech and language processing* (Vol. 3). London: Pearson.

Kauhanen, H. 2017. Neutral change *Journal of Linguistics*, 53(2), 327-358.

Kershaw, D., Rowe, M. and Stacey, P. 2016, February. Towards modelling language innovation acceptance in online social networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* (pp. 553-562). ACM.

Labov, W. 1966. The linguistic variable as structural unit. Wash. *Linguist* Rev. 3, 4–2*2.

Maybaum, R. 2013, December. Language change as a social process: Diffusion patterns of lexical innovations in Twitter. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 39, No. 1, pp. 152-166).

Metcalf, A. 2004. *Predicting New Words. The Secrets of Their Success* Houghton Mifflin Harcourt.

Milroy, J. and Milroy, L. 1985. Linguistic change, social network and speaker innovation. *Journal of linguistics*, 21(2), pp.339-384.

Naili, M., Chaibi, A.H. and Ghezala, H.H.B. 2017. Comparative study of word embedding methods in topic segmentation. *Procedia computer science*, 112, pp.340-349.

Nation, P. and Waring, R. 1997. Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, 14, pp.6-19.

Nevalainen, T. and Raumolin-Brunberg, H. 2016. *Historical sociolinguistics: language change in Tudor and Stuart England*. Routledge.

Phillips, B. 2006. *Word frequency and lexical diffusion*. Springer.

Piantadosi, S.T. 2014. Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, 21(5), pp.1112-1130.

Rogers, Everett M. 1962. *Diffusion of innovations*. New York: Free Press of Glencoe.

Sornig, K. 1981. *Lexical innovation: A study of slang, colloquialisms and casual speech*. John Benjamins Publishing.

Tamburrini, N., Cinnirella, M., Jansen, V.A. and Bryden, J. 2015. Twitter users change word usage according to conversation-partner social identity. *Social Networks* 40, pp.84-89.

Thomsen, O.N. ed. 2006. *Competing models of linguistic change: evolution and beyond* (Vol. 279). John Benjamins Publishing.

Trudgill, P. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in society*, 3(2), pp.215-246.

Virk, Amardeep. 2011. Twitter: The strength of weak ties. *University of Auckland Business Review*, 13(1), p.19.

Yamanouchi, Y, and Komatsu, T. 2014. Power Law in SNS Language Probability Space and Stable Distribution. Japan: In *Journal of The Infosocionomics Society* (vol. 11, No. 1)

Yang, C.D. 2000. Internal and external forces in language change. *Language variation and change*, 12(3), pp.231-250.