Emerging English Transitives over the Last Two Centuries

Liwen Hou
Khoury College of Computer Sciences
Northeastern University
Boston, MA
hou.l@northeastern.edu

David A. Smith
Khoury College of Computer Sciences
Northeastern University
Boston, MA
dasmith@ccs.neu.edu

Abstract

We analyze an automatically-parsed British Hansard to identify approximately 200 verbs that first appeared in transitive constructions in British English in the 19th and 20th centuries. We use this list of verbs to test two hypotheses about new verb forms. First, we test the hypothesis that rarer verb lemmas are more likely to experience language change compared to more common verb lemmas. As measured by our specific notion of language change, we find that this is true only up to a certain rarity, and extremely rare lemmas are actually less likely to change compared to somewhat rare lemmas. Second, for new transitive verbs, we test the hypothesis that the passive construction is introduced later than its active counterpart. We find some evidence for this hypothesis but show that it is not universally true by exhibiting several verbs whose active and passive usages were introduced simultaneously.

1 Introduction

How are new verbs introduced? Under what circumstances do intransitive verbs become transitive? In this work, we set out to investigate emerging transitive verbs using automated methods. Primarily, we identify approximately 200 new transitive verbs that were introduced in the 19th and 20th centuries (after the year 1849) in the record of British parliamentary debates. We use this dataset, commonly known as the British Hansard, to test two hypotheses regarding new transitive verb formations: One hypothesis, which we call the frequency hypothesis, says that rare words change more readily than common words. In our context, this would also imply that rare nouns become verbs more readily than frequently occurring nouns. The second hypothesis is that the passive forms of new transitive verbs emerge later than their active counterparts on average. Note that transitive verbs that are new to the English language can come from several sources: Some (e.g. "collate") used to occur as intransitive verbs. Some (e.g. "chair") previously appeared as nouns. Others (e.g. "highlight") did not exist at all at the beginning of the 19th century.

2 Related Work

Previous work using computational approaches to answer questions about language change has focused on a variety of tasks, the most common of which are the following: inferring linguistic phylogenies, identifying cognates, reconstructing proto-languages, tracking word senses and semantic change over time, and understanding morphological change.

Establishing relationships between different languages through the reconstruction of linguistic phylogenies is the focus of several lines of work (Gray and Atkinson, 2003; Gray et al., 2009; Nakhleh et al., 2005). Other papers focus on the cognate identification task (Kondrak, 2001; Mackay and Kondrak, 2005; Hall and Klein, 2010). The reconstruction of proto-languages is another historical linguistics task for which computer scientists have proposed probabilistic methods; examples of work in this area include Bouchard-Côté et al. (2008) and Bouchard-Côté et al. (2013).

In addition, there have been many attempts to detect lexical semantic change using word embeddings. Works such as Kim et al. (2014), Kulkarni et al. (2015), and Hamilton et al. (2016) use this approach to identify changes in lexical categories and word senses (e.g. "mouse" gained another meaning as computers became more widespread).

Focusing on morphological change, Lieberman et al. (2007) considers English verbs from the last 1200 years and examines the process by which these verbs have become regular. They conclude that the half-life of an irregular verb is correlated

Proceedings of the Society for Computation in Linguistics (SCiL) 2021, pages 71-80.
Held on-line February 14-19, 2021

with its usage frequency and that "a verb that is 100 times less frequent regularizes 10 times as fast."

Another related work in morphology is Kisselew et al. (2016), focusing on nouns that changed to become verbs as well as verbs that became nouns; their goal was predicting the form that appeared first.

Although the task of automatically detecting syntactic change has not received much attention in the literature, syntactic change detection using a single-layer LSTM is the topic of Merrill et al. (2019). Because the LSTM achieved higher accuracy for the task of predicting the year of composition of novel sentences compared to their baseline model consisting of feedforward networks, and because the latter do not take relations between words into account whereas LSTMs do, Merrill et al. (2019) conclude that syntactic change (as opposed to only lexical change) was detected.

3 Data and Methods

3.1 Hansard Corpus

The dataset used in our analyses is the British Hansard corpus, which consists of over seven million speeches from the U.K. Parliament spanning over two centuries. A digitized version of the debates from 1803 to 2003 can be found on the U.K. Parliament Website's Hansard archive[1].

3.2 Classifying Verbs

We obtain labeled dependency parses and part-of-speech tags using Stanford CoreNLP's syntactic dependency parser and POS tagger (Manning et al., 2014; Chen and Manning, 2014; Toutanova et al., 2003). For every instance of every verb, we then attempt to determine whether or not it takes a direct object using the parser output. To accomplish this, our approach is as follows.

We consider one sentence at a time (discarding sentences that appear more than twice in a parsed file), processing from the top of the file downwards using primarily the dependency relations produced by the parser. For all *dobj* dependency relations, we lemmatize the head word if we have not yet considered it before (as we want to avoiding double counting and concluding that the same token is associated with more than one dependency relation), then we count this as one instance of the lemmatized head word being used as a direct object. In addition to the relations identified as *dobj* by the parser, we create a new *dobj* relation whenever two verbs are coordinated and the first of the two verbs appears with a direct object, since the parser fails to identify the *dobj* relation that exists between the second verb and the object. For example, when given the sentence in (1), the parser outputs a *dobj* relation between *stolen* and *documents* but no other *dobj* relation (even though there are two verbs that take *documents* as a direct object in this sentence).

(1) The way they have stolen and leaked documents is quite disgraceful.

Therefore, we add a *dobj* dependency relation between *leaked* and *documents* to our parsed data in this case.

Next, if the dependency relation is *auxpass* or *nsubjpass*, we lemmatize the head word and classify that instance as a passive usage (as long as the head word has not already been classified). If ever the same word appears as the head word of more than one dependency relation, it will be given a classification only the first time.

If the dependency relation is *npadvmod*, then we ignore the head word, since this relation often captures noun phrases that are inside of a VP but that are not real objects, as in "shares eased a fraction". In addition, if the head word is a noun while the dependent is a verb, then we ignore the dependent when the relation is not *rcmod*; that is to say, unless a verb is the head of a relative clause, we ignore that verb if it is governed by a noun, so that we avoid dealing with NPs such as *ticketing machines*.

At this point, if any word in the sentence was tagged as a verb but has not yet been classified, then we lemmatize it and consider that instance of the word as not having a direct object.

We do some further cleanup to mitigate the errors made by the parser and POS tagger: We remove any words with special characters (other than hyphens) from the list of identified verbs, any words with fewer than three characters, as well as any words that lack vowels (as the latter are usually abbreviations).

In addition, from the list of verbs, we remove any lemma that occurs as a verb less than 10% of the time. The reason we do this is that the POS tagger sometimes makes errors, and this error rate is amplified when we focus on an unusual usage of a word. For example, consider a lemma that is used as a noun 99% of the time and as a verb 1% of the time, such as "book". Suppose the POS tagger misclassifies 1% of the noun occurrences of "book"

[1] http://www.hansard-archive.parliament.uk/

as verbs, and suppose it correctly classifies the verb occurrences of "book" as verbs. Then among the instances of "book" that have been classified as verbs, only 50% are correct. This means that a POS tagger with an error rate of 1% is not accurate enough to allow us to study rare verbs such as "book". To combat this issue, we restrict the verbs under consideration to only those whose lemma occurs as a verb at least 10% of the time.

Finally, we merge hyphenated words with their unhyphenated counterparts if both exist, and we display the word as the more common variant.

4 Identifying Change

4.1 Chi-squared Test

In order to identify verbs that have undergone change, we first bin the data into an early bin (spanning the period from 1800 to 1849) and a late bin (spanning the years from 2000 to 2003); we leave the 150 years from 1850 to 1999 outside of both bins. Note that the number of words spoken per year increases significantly over time in our dataset; in the period from 1800 to 1849, our dataset contains 51.6 million words, while in the period from 2000 to 2003, our dataset contains 66.3 million words. Therefore, although our bins cover vastly different amounts of time, they include a similar number of words.

In each bin, and for each lemma, we count the number of times the lemma occurred in the bin as a transitive verb. We use these counts, together with the total size of the bins, to draw a contingency table, such as the one in Table 1 for the lemma "eat".

	Earlier period	Later period
transitive "eat" count	164	253
count of other lemmas	51560849	66258950

Table 1: Contingency table for "eat".

Using this contingency table, we then apply a χ^2 test for statistical significance. For this significance test, the null hypothesis is that the rate of occurrence of the transitive verb "leak" in the earlier period is equal to its rate of occurrence in the later period.

We apply the χ^2 test to each transitive verb that had more than 10 total occurrences over both the earlier and later bins. There are 2864 such transitive verbs in total. We set a p-value threshold that is equal to $0.01/2864 \approx 3.5 \times 10^{-6}$, which we obtain by choosing a significance threshold of 0.01 and doing a Bonferroni adjustment for multiple comparisons. The result of this test is that 1439 transitive verbs have a p-value smaller than this significance threshold, and hence have had a statistically significant change in usage over the period 1850-1999. Note that although the changes are statistically significant, some are small in magnitude. The change may also be a consequence of changing topics in parliamentary debates rather than true language change.

4.2 Identifying New Transitive Verbs

According to Visser (1963), p. 99, "the most remarkable fact is that, whereas in Old English the number of double-faced or amphibious verbs was far inferior to that of intransitive verbs, in Pres. D. English amphibious verbs far outnumber the intransitive verbs" (where the author's meaning of an *amphibious* verb is one that is sometimes transitive and other times intransitive). Visser noted that, although this process of becoming transitive started in Old English, it also affected the intransitive verbs that first appeared later on. Visser then proceeded to listing 58 intransitive verbs from Present-Day English and noted that a perusal of the Oxford English Dictionary would not show many more.

Given that intransitive verbs have become quite rare in the English language, we choose to focus on identifying new transitive verbs in this section.

We restrict our attention to transitive verbs that occurred at least 20 times in the final years of the dataset (i.e. from 2000 to 2003) but that did not occur at all in the first 50 years of the data (i.e. from 1800 to 1849). We note that the χ^2 test for a verb that occurred 0 times in the early period and 20 times in the late period gives a p-value of 0.0002, so all the changed verbs we identify are significant to at least this level (sometimes much more).

This yields 196 new transitive verbs[2] (see Table 2), and we observe that even after a Bonferroni correction for multiple comparisons, their changes are all significant at the $p < 0.05$ level (and indeed, the more common ones are significant to $p < 10^{-200}$). Sorting by the frequency of the verb in the 2000-2003 period, the top 10 new transitive verbs are highlight, monitor, outline, chair, clarify, contact, target, underpin, envisage, and stress; of these, the first seven occurred over 1,000 times in the 2000-2003 period, which means that they each occurred

[2]The count of 196 does not include "outwith", "criteria", "broadband", and "trial", which originally also qualified for Table 2 but only due to errors in lemmatization or POS tagging.

DECADE									
1850s	revolutionise	legitimise	overhaul	permeate	condone	shoulder			
1860s	eliminate	manipulate	intensify	recoup	skip				
1870s	underlie	formulate	dilute						
1880s	deplete	accentuate	differentiate	chuck	trail	interview	tip	champion	safeguard
1890s	summarise	resurrect	dominate	minimise	circumvent	forgo	diagnose	penalise	belittle
	utilise	mobilise	typify	flout	earmark	grab	undercut	forecast	raid
	voice	pool							
1900s	broaden	centralise	delete	crystallise	terrorise	re-read	dodge		
1910s	deflect	standardise	stabilise	symbolise	exploit	scrap	bomb	guillotine	grade
1920s	harmonise	equate	castigate	dump	stress	broadcast	trap	chart	tour
	sample	park							
1930s	envisage	rationalise	blur	sidestep	outline	co-ordinate	query	offset	
1940s	bedevil	devalue	revitalise	integrate	clarify	truncate	enthuse	activate	epitomise
	denigrate	slash	re-emphasise	redirect	reassess	tidy	sponsor	swap	bypass
1950s	modernise	inject	erode	unleash	disrupt	motivate	streamline	sadden	publicise
	validate	deploy	align	redefine	rewrite	claw	tailor	plug	sub-contract
1960s	categorise	rephrase	peddle	exacerbate	quantify	extradite	formalise	renegotiate	straddle
	abut	downgrade	rethink	fiddle	alert	leak	span	phase	ditch
	breach	spot							
1970s	underscore	evaluate	toughen	finalise	donate	refurbish	subsume	pre-empt	encapsulate
	uprate	scupper	hijack	disapply	over-egg	predate	highlight	update	rumble
	restructure	focus	trigger	boost	muddy	upgrade	duck	chair	lobby
	contact	spearhead							
1980s	destabilise	rubbish	disaggregate	privatise	reinvent	short-change	underfund	monitor	outperform
	decommission	trumpet	co-sponsor	complement	log	scar	buck	phone	
1990s	collate	replicate	criminalise	prioritise	cherry-pick	underpin	ring-fence	recycle	refocus
	tick	second-guess	host	fax	target				
2000s	incentivise	outsource							

Table 2: New transitive verbs that emerged after the mid-19th century (grouped according to the decade in which the transitive usage reached 20% of its final usage frequency)

with a frequency of more than 1 in every 66,300 words.

Note that some of these words, such as “highlight”, “envisage”, and “underpin”, are new words that did not exist in any form in the 1800-1849 period. Other words, such as “chair”, “stress”, and “outline”, occurred as nouns. Some words, such as “collate”, occurred in the 1800-1849 period as intransitive verbs only (in the case of “collate”, it was used in the period from 1800 to 1849 in the sense of “appoint” instead of the modern sense meaning “collect” or “combine”).

5 Testing Hypotheses

Next, we examine two classes of hypotheses from the diachronic literature on English.

5.1 Frequency Effects

We analyze the effect of frequency on the tendency of words to adopt a new transitive usage. For syntactic change, Bybee and Thompson (1997) noted that “on the one hand, high token frequency promotes change and on the other hand it renders constructions resistant to change”. For semantic change, Hamilton et al. (2016) found that the meanings of more common words are more stable, while uncommon words gain new meanings more easily. For morphological change, Lieberman et al. (2007) found that rarer verbs regularized faster.

For this paper, the frequency hypothesis is therefore that more frequent words change more slowly; specifically, we ask whether more frequent words are less likely to add a new transitive usage.

In our current context, we group all the lemmas used in the 50-year period from 1800 to 1849 by frequency, to examine whether the less frequent ones are more likely to add a transitive use case. Recall that we identify new transitive use cases for which the transitive verb form occurred at least 20 times in our dataset in the years ranging from 2000 to 2003, which approximately corresponds to the new transitive verb form occurring at least once in every 3 million words by the beginning of the 21st century. New verb forms that occur significantly more rarely than this cannot be identified given the size of our dataset.

After lemmatizing the words spoken from 1800 to 1849, we bin them by the number of times they occurred, with each bin containing words whose counts are within a multiplicative factor of 3 of each other. That is, the first bin might contain the words that occurred 1 to 3 times, the second might contain the words that occurred 4 to 9 times, the third might contain the words that occurred 10 to 27 times, and so on. More formally, we take the logarithm (base 3) of the number of occurrences of each lemma, and we then round the result to the nearest integer; this integer is considered to be the frequency bin of that lemma.

Within each frequency bin, we then consider the fraction of lemmas that added a new transitive use case. This is simply the total number of lemmas in the bin that are in the set of new transitive verbs we have discovered, divided by the total number of lemmas in the bin. The interpretation of this is as follows: if someone from the year 1850 wanted to know the likelihood of a given lemma to later become a transitive verb, that person should take the frequency of that lemma and go to the corresponding frequency bin; the likelihood is then the fraction of the lemmas in that bin that newly became transitive verbs by the beginning of the 21st century. We plot these numbers in Figure 1.

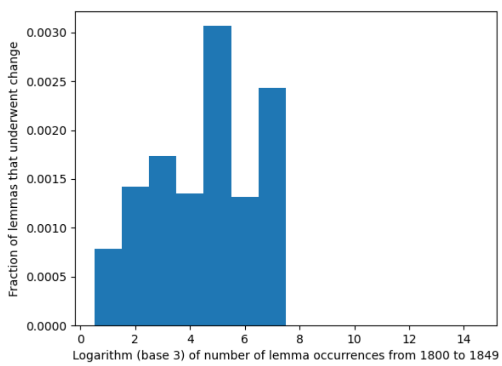


Figure 1: Probability of becoming a transitive verb, by log frequency. Each bin i corresponds to lemmas that occur with frequency roughly $\frac{3^i}{5\times10^7}$. The height of a bin is the ratio between the number of lemmas that newly became transitive verbs and the total number of lemmas of that frequency. Bin 7 has only 3 lemmas that became transitive verbs and is thus particularly noisy.

As the figure shows, common verb lemmas (i.e. those occurring with a frequency of 1 in 10,000 or more) are not among the lemmas we found to have newly become transitive verbs. However, extremely rare lemmas, such as those occurring with a frequency of 1 in 10,000,000 in our dataset, are also unlikely to add a new transitive verb form (or at least, a new verb form that occurs with sufficient probability for us to identify it as such). This is in part because there are simply so many rare lemmas that even though several of them do become transitive verbs, the probability of any given one of them experiencing such a change is small.

The peak change seems to occur in lemmas that have a frequency of approximately 1 in 100,000; such lemmas include "stress", "shoulder", and "interview" (each of which newly became a transitive verb in our dataset).

The existence of a peak frequency in which language change happens is in contrast to other work on frequency effects for language change, which mostly find that less frequent words change more readily (with no peak at any frequency, or an implicit peak at 0). One potential reason for this change is that we are asking a somewhat different question. Instead of asking how fast each word changes over time, we ask how likely each word is to exhibit a significant change, and we only count a change as significant if the number of occurrences of the new form in the later period is sufficiently large. While there is nothing in principle that prevents a rare word from adopting a new meaning that is commonly used, such instances are necessarily somewhat uncommon: there are only so many words that can have new common uses in the later time period, while there are a very large number of very rare words in the early time period. In this light, the existence of a peak frequency for this type of language change is perhaps less surprising.

5.2 Delayed Passivization Effect

Christiansen and Joseph (2016) observed that passivization is usually the test for direct object status, and they also observed that there exist cases of verb phrases that fail the passivization test even in the absence of a preposition; for example:

(2) Jordan rocked.

(3) Jordan rocked that blouse.

(4) *That blouse was rocked by Jordan.

The verb in (2) is intransitive. The verb in (3) is transitive. The verb in (4) is the ungrammatical passive corresponding to the meaning in (3).

The above three examples from Christiansen and Joseph (2016) suggest that newly transitive verbs sometimes do not sound natural in the passive voice. This observation leads us to hypothesize that, for new transitive verbs, the passive construction emerges later than the active on average.

To investigate this hypothesis, we use the new transitive verbs that we identified. For each one, we use the parser to classify the instances of its use as active or passive. Note that Hou and Smith (2018) found that the Stanford parser could correctly identify the voice of a verb in the majority of cases after a manual evaluation of the accuracy of the parser.

It is then possible to find the first year in which the active and passive forms were used and to compare them to each other. Doing so, we obtain the result that, on average, the first active instance of the new transitive verb occurs 33 years before the first passive instance. However, there are several issues with this naive analysis. First, it hides a large amount of variance between the different new verbs. Second, the first instance is not necessarily a good measure of when the verb truly started being used, as it is sometimes the case that a verb appears just once several decades before it starts appearing regularly (possibly due to errors in the dataset). A third problem with this approach is that the background rate of passive constructions is lower than that of active constructions (even for verbs that did not recently change); this means that even if the new transitive verbs had the exact same passivization patterns as the old verbs, we would expect the first time a new transitive verb is used to be in the active voice, simply because the active is more common overall.

We try to address some of these issues. We first check whether, at the end of our dataset, the passivization rate of the new transitive verbs patterns differently from the passivization rate of old transitive verbs. To this end, we plot histograms of the passivization rates of the new and overall transitive verbs in Figure 2 in order to visualize the difference.

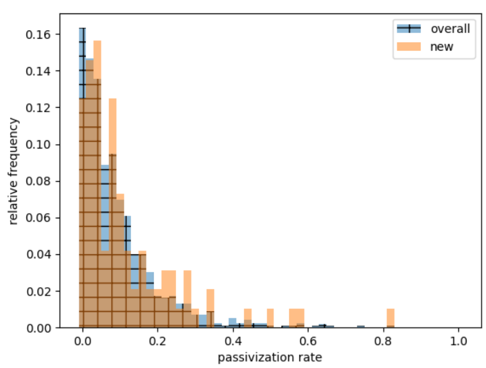


Figure 2: Histograms of the passivization rates of the new and overall transitive verbs in the years from 2000 to 2003. The x-axis is the passivization rate of verbs, and the y-axis is the relative frequency of that passivization rate among the set of verbs.

As the figure shows, the passivization rates of the new transitive verbs look very similar to the passivization rates of transitive verbs overall by the beginning of the 21$^{\text{st}}$ century. This rules out a long-lasting delay in passivization for new transitive verbs, but it does not rule out more modest delays, as it is conceivable that although the passive construction is delayed, it already exists by the end of our dataset.

Another approach to measuring the passivization delay for new transitive verbs is to look at the first time the frequency of the active form of the new verb exceeds some threshold and to compare this to the first time the frequency of the passive form exceeds some threshold. We pick the threshold for the active form to be 20% of the frequency of the active form of verb at the end of the dataset, and similarly we pick the threshold for the passive form to be 20% of the frequency of the passive form of the verb at the end of the dataset.

The above comparison leads to the conclusion that the passive construction lags behind the active construction by an average of 7 years. However, this number once again hides a lot of variation. Moreover, using a threshold that is 20% of the final rate gives misleading results in the case where the frequency of the passive form of a verb is stable but low for a long period of time, while the active form keeps increasing; for such verbs, the passive will reach 20% of its final value very early on, while the active will not reach 20% of its final value until much later (since the final frequency of the active construction is much higher).

5.3 Logistic Growth Model

Next, we describe a model for characterizing the change in frequency over time. For every pair consisting of a verb and a usage category (e.g. passive), our model first generates a rate of usages for that subcategorization frame at every time period. We observe the number of relevant occurrences, which are counts that are distributed according to a Poisson distribution whose underlying rate follows a logistic function of time.

The number of occurrences for a given pair (consisting of a verb and a usage category) is generated by a Poisson process. For all t ranging from 1 to T, we have the count at time t being generated by

$$\text{poisson}\left(n_t \cdot \left(er + \frac{lr - er}{1 + e^{-k(t-p)}}\right)\right)$$

where n_t is the amount of data that we have for time t, k is the magnitude of the slope, er is the earlier rate, lr is the later rate, and p is the inflection point.

The inflection point is the midpoint of the change and also where the slope is the steepest (i.e. where the change is happening the fastest). The earlier rate occurs at time negative infinity, and the later rate occurs at time infinity. (The transition between the two occurs with a slope of magnitude k.)

Since we are only looking at verbs that did not exist in the early period, we set $er = 0$. For the remaining parameters, we use the following priors:

$$\begin{aligned} p &\sim \text{uniform}(0, T) \\ k &\sim \text{exponential}(r_k) \\ lr &\sim \text{exponential}(r_l) \end{aligned}$$

For our dataset, T is the number of years, which is 203. We set $r_k = 0.1$ and $r_l = 10^{-5}$. We fit this model using the Stan language (Carpenter et al., 2017; Stan Development Team, 2018).

This model has the following limitations: If a particular usage rises and then subsequently falls (or vice versa), the logistic curve will not be able to capture the change. Additionally, increases that occur in multiple discrete steps, or which otherwise do not look like a logistic curve, will not be well-characterized by this model.

5.4 Applying the Model

The logistic curve is a good fit for characterizing the change in some of the verbs, though not all of them. In Figure 3, we provide some plots of the active and passive frequencies and their associated fits obtained from our model.

To determine whether the active or the passive usage appeared first, we first restrict our attention only to those verbs in which the final fitted rate lr for the passives is at least 20% of the final fitted rate lr for the actives. We do this in order to avoid verbs for which the active usage increases sharply over time but for which the passive usage does not; for these verbs, it is not clear whether we should consider the passive construction to be delayed beyond the end of our dataset, or whether we should consider the passive construction to have completed the change already (albeit at a very low final rate of usage). We also eliminate verbs for which the mean squared error of either the active or passive logistic curves is too high, indicating a poor fit, as well as verbs for which the parameter k is too small, indicating that the detected change is too gradual and slight to meaningfully estimate the inflection point p.

For the remaining verbs, we compare the inflection point p of the active form to the inflection point of the passive form. On average, the inflection point of the active form of a new verb occurs 8 years before the inflection point of the passive form, indicating a potential small delay in passivization for new verbs. The variance is once again fairly large; the difference between the passive inflection point and the active inflection point of a verb ranges from -7 to 52 years (for the verbs “chair” and “offset” respectively).

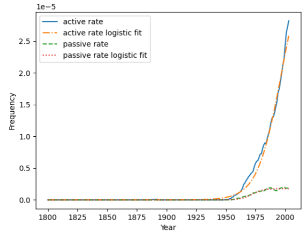
5.5 Discussion

This comparison of inflection points is still not a perfect way to measure a potential delay in passivization for new verbs. One issue is that when the change is gradual, the inflection point occurs much later than the time when the curve starts to rise. A second issue is that when the rate of usage of a verb form is still increasing at the end of our dataset (instead of leveling off), the estimate of the inflection point is not very stable.

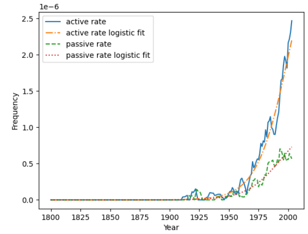
Therefore, as a final check, we manually examine the plots for all new transitive verbs to see whether a delayed passivization effect seemed apparent. From a manual examination, it appears that some verbs have the active and passive uses increase together at the same time, while for other verbs the passive remains nearly nonexistent throughout the period of time covered by dataset (even as the active form rises in frequency). While this gives some evidence towards a delayed passivization effect, it could also be that some transitive verbs are rarely used in the passive voice for reasons other than being new. One notable finding, however, is that there are no verbs for which the passive form was unambiguously introduced before the active form (while the reverse occurs).

6 Conclusion

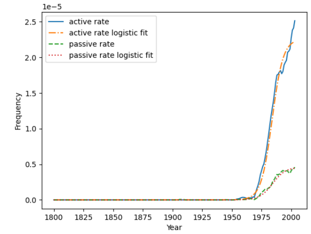
We automatically identified new transitive verbs that first appeared in British English after the mid-19$^{\text{th}}$ century. We used these verbs to test two hypotheses about new verbs. First, we tested the hypothesis that rarer verb lemmas are more likely to undergo change compared to more common lemmas; we found this to be true only up to a certain rarity, as extremely rare lemmas are actually less likely to undergo change compared to somewhat rare lemmas. Second, we tested the hypothesis that the passive construction emerges later on average



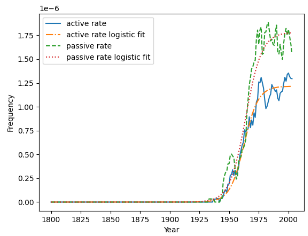
(a) Highlight



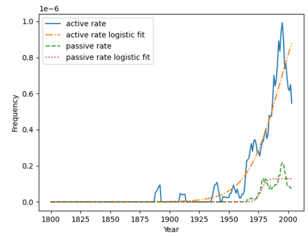
(b) Lobby



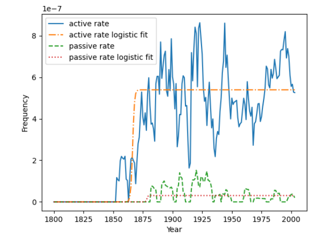
(c) Monitor



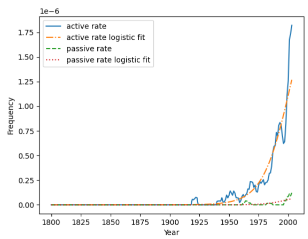
(d) Motivate



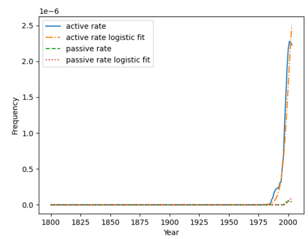
(e) Peddle



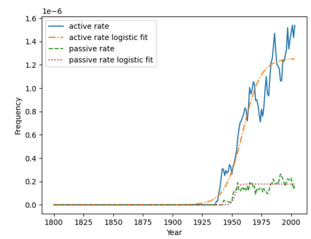
(f) Permeate



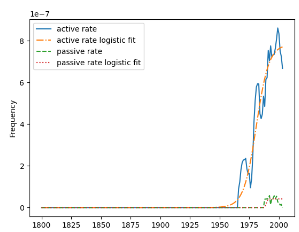
(g) Phone



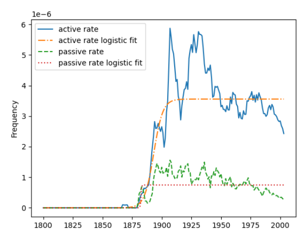
(h) Prioritise



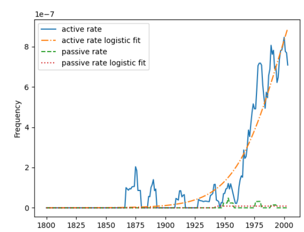
(i) Publicise



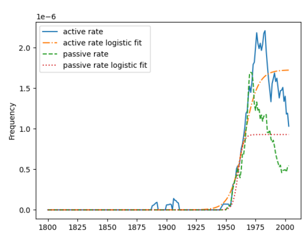
(j) Restructure



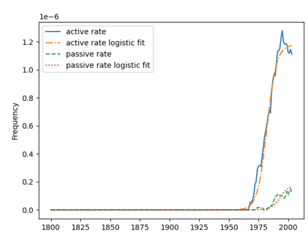
(k) Safeguard



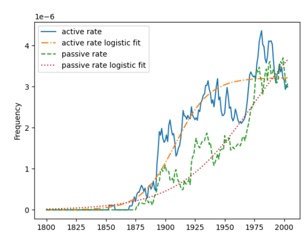
(l) Span



(m) Erode



(n) Encapsulate



(o) Dominate

Figure 3: Logistic fits for the active and passive frequencies over time of selected new transitive verbs, where the active and passive frequencies are drawn as 5-year rolling averages.

compared to its active counterpart; while we were able to present some evidence for this hypothesis, we also found that it is not always true, and we showed some verbs whose active and passive usages were introduced at approximately the same time. Finally, we described a logistic growth model and provided plots of the active and passive frequencies over time of several new verbs along with their associated fits obtained from our model.

Possible directions for future work include separately examining different types of new transitive verbs that we grouped together in this work: intransitive verbs that developed a transitive usage over time, nouns that became verbs, and brand new words. It would also be interesting to study similar phenomena using a language other than English, especially if the language under consideration has seen as many transitive verbs become intransitive over time as the other way around; English does not have this property, as verbs have rarely become intransitive in the history of English (Visser, 1963).

Acknowledgments

We are grateful to Janet Randall for the discussions on argument structure. We also thank the anonymous reviewers for their comments.

References

Alexandre Bouchard-Côté, David Hall, Thomas L Griffiths, and Dan Klein. 2013. Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences*, 110(11):4224–4229.

Alexandre Bouchard-Côté, Percy S Liang, Dan Klein, and Thomas L Griffiths. 2008. A probabilistic approach to language change. In *Advances in Neural Information Processing Systems*, pages 169–176.

Joan Bybee and Sandra Thompson. 1997. Three frequency effects in syntax. In *Annual Meeting of the Berkeley Linguistics Society*, volume 23, pages 378–388.

Bob Carpenter, Daniel Lee, Marcus A Brubaker, Allen Riddell, Andrew Gelman, Ben Goodrich, Jiqiang Guo, Matt Hoffman, Michael Betancourt, and Peter Li. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Danqi Chen and Christopher D Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 740–750.

Bethany J Christiansen and Brian D Joseph. 2016. On the relationship between argument structure change and semantic change. *Proceedings of the Linguistic Society of America*, 1(26):1–11.

Russell D Gray and Quentin D Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426(6965):435–439.

Russell D Gray, Alexei J Drummond, and Simon J Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323(5913):479–483.

David Hall and Dan Klein. 2010. Finding cognate groups using phylogenies. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1030–1039, Uppsala, Sweden. Association for Computational Linguistics.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501.

Liwen Hou and David A Smith. 2018. Modeling the decline in English passivization. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2018*, pages 34–43.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal analysis of language through neural language models. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 61–65.

Max Kisselew, Laura Rimell, Alexis Palmer, and Sebastian Padó. 2016. Predicting the direction of derivation in English conversion. *Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 93—98.

Grzegorz Kondrak. 2001. Identifying cognates by phonetic and semantic similarity. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web*, pages 625–635.

Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang, and Martin A Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature*, 449(7163):713–716.

Wesley Mackay and Grzegorz Kondrak. 2005. Computing word similarity and identifying cognates with pair hidden Markov models. In *Proceedings of the*

Ninth Conference on Computational Natural Language Learning (CoNLL-2005), pages 40–47.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David 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, Baltimore, Maryland. Association for Computational Linguistics.

William Merrill, Gigi Felice Stark, and Robert Frank. 2019. Detecting syntactic change using a neural part-of-speech tagger. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 167–174.

Luay Nakhleh, Don Ringe, and Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language*, pages 382–420.

Stan Development Team. 2018. Pystan: the Python interface to Stan, version 2.17.1.0. `http://mc-stan.org`.

Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology-volume 1*, pages 173–180. Association for Computational Linguistics.

Frederik Theodor Visser. 1963. *An Historical Syntax of the English Language: Part One, Syntactical Units with One Verb*. EJ Brill.