CS224N/LING284 Final Project:
Measuring Functional Load with Word Vectors

Simon Todd
sjtodd@stanford.edu

1 Introduction

Languages are constantly changing – slowly but surely. The shape of a language in the present day is a reflection of its shape in the past and of processes that have affected it over time. Thus, a fundamental component of understanding the system of language is understanding the forces behind these historical processes. Furthermore, given such an understanding and a representation of the state of a language today, it is possible to make predictions about how the language is likely to change in the future. It is for these reasons that the study of language change is interesting.

Numerous hypotheses about the forces behind language change have been proposed. Here, I present a pilot study focusing on one particular hypothesis about the way that sounds in a language change over time: the Functional Load Hypothesis. I replicate past corpus studies investigating the hypothesis with new data and new methods: in particular, I investigate the potential use of word vector similarity as a simple, low-needs metric for evaluating the hypothesis.

2 Background

2.1 Phoneme systems and sound change

The sound system of a language is composed of **phonemes**: sounds which create differences in meaning. For example, the sounds [p] and [t] correspond to different phonemes (notated /p/ and /t/ respectively) because the words *pin* and *tin*, which are identical save for those sounds (i.e. form a **minimal pair**), have different meanings[1]. However, over time the sound system of a language changes and some of these contrasts are lost, a phenomenon known as **merger**. For example, the vowels in the words *cot* and *caught* (/ɒ/ and /ɔ/ respectively) are indistinguishable for some General American English speakers, though they were distinct when American English split off from British English. Though the number of mergers within any one language system is small, the phenomenon of merger is widely attested cross-linguistically and across time; in other words, it appears to be a robust component of language dynamics. The question therefore arises: can we predict which phonemes in a language are likely to merge and which are likely to remain distinct?

2.2 The Functional Load Hypothesis

It has been hypothesized since at least Martinet [1] that a major factor in determining whether or not two phonemes in a given language will merge is their **functional load**; that is, how much work is uniquely done by the phonological contrast in keeping apart (words with) different meanings in the language. Recently, Wedel et al. have shown that the predictions of the Functional Load Hypothesis are quantitatively borne out cross-linguistically: using corpora from 9 different languages (British RP English, General American English, Korean, French, German, Dutch, Slovak, Spanish, and Hong Kong Cantonese), they show that the greater the number of words distinguished by a given phoneme pair in a language, the lower the probability that pair of phonemes will merge [2], and that the strength of this relationship is increased if the minimal pair counts are weighted by word frequency ratio and if only contrasts between words that have the same POS are considered [3].

The intuition behind these results is simple. There are many factors which may act to disambiguate homophonous (same-sounding) word-pairs in language use. For example, one word may be impossible because it does not fit with the requirements of the syntactic context; for speakers of American English with the COT $\sim$ CAUGHT merger, the word *cot* will never be mistaken for the word *caught* in the context of the word *the*,

[1] The symbols used in this paper are those of the International Phonetic Alphabet

as verbs cannot immediately follow determiners in English (but nouns can). Similarly, the word *bodily* will never be mistaken for the word *bawdily* because the latter is highly infrequent and thus highly unlikely *a priori*. When these factors are not at play – when the two words in a minimal pair have the same part of speech or are similar in frequency – the danger of confusion arising from homophony is high. If language is functional, i.e. sensitive to the needs of efficient and robust information transfer, then such confusion should be avoided. Thus, if neutralization of a given phoneme contrast through merger would introduce homophonous word-pairs which cannot readily be disambiguated by other means, then that contrast carries the load of keeping the words apart and should not be lost.

2.3 Proposal

Wedel et al.'s result as concerns the importance of within-POS contrasts is intriguing, but limited, both practically and theoretically. Practically, because their measure required knowing the POS tag of every word in a corpus, it is difficult to extend to languages where a corpus and a lexicon are available but not (reliably) POS-tagged. Theoretically, while two words with the same POS are likely to be more substitutable and thus more confusable than two words with different POSs, putting more load on the phonology for keeping the two words apart, the notion of "within POS category" is very broad and there are still many instances where two words with the same POS are readily differentiated from context (lowering the functional load on the phonological contrast). What is needed is a metric for functional load which (a) can be computed from unannotated data, and (b) more readily reflects how substitutable two words are.

I propose that the cosine similarity of word vectors could constitute such a metric. Word vectors capture the distributional properties of words. Thus, two words will have similar word vectors if they are often used in the same environment. Intuitively, it is very important for words which can appear in the same environment to be distinguished in some way; without such distinction, it would be incredibly difficult for a listener to identify which word was meant by a speaker. Since minimal pairs are only distinguished in form by a phonological contrast, it follows from the Functional Load Hypothesis that members of a minimal pair which have highly similar distributions (word vectors) will place a lot of load on that contrast and provide strong barriers to its loss, while members of a minimal pair whose distributions are almost non-overlapping will place almost no load on the contrast and will not provide barriers to its loss (since the meanings they signal can be identified robustly from context).

3 Methods

I present a pilot study testing the applicability of word vector similarity to measurements of functional load. In addition, I replicate Wedel et al.'s results for different measurements of functional load on larger datasets drawn from 2 of the 9 languages they investigated.

3.1 General framework

Let $\mathcal{L}_X$ be a phonologically-annotated lexicon of language X; i.e. a function from orthographic strings o to phonological strings p representing words in X. The phonological strings are sequences of phonemes: $p = p_1 p_2 p_3 ... p_n$.

Let $\mathcal{M}_X^C$ be the set of minimal pairs in the language with respect to the phonological contrast C, which is a triple (α, β, e) composed of two phonemes α and β and a phonological environment e. Two separate orthographic strings o, o' in the language are minimal pairs with respect to C if their phonological forms p and p' respectively are identical, except that one has α whenever the other has β in the environment e.

Let W be a function which applies a weighting to each minimal pair $(o, o') \in \mathcal{M}_X^C$ in some way. Then the **functional load** F_X^C of the contrast C in the language X can be defined as in Equation (1):

$$F_X^C = \sum_{(o,o') \in \mathcal{M}_X^C} W(o, o') \qquad (1)$$

Thus, the functional load of a phonological contrast is simply a weighted sum of the minimal pairs with respect to that contrast.

I explore 5 different weighting functions, constituting different measurements of functional load, defined in Sections 3.3 and 3.4.

3.2 Data

I investigated two varieties of English: British RP English and General American English. Within these languages, I considered all phonological contrasts whose constituent phonemes differ in a single feature such as voicing or place of articulation, across several different subsystems. This yielded a greater number of unmerged contrasts than was investigated by Wedel et al.. Following Wedel et al., 10 phonological contrasts in each language were coded as merged; the full list of contrasts is given in Table 2 (at the end of the document).

I obtained a lexicon for each of these languages from the Unisyn speech synthesis lexicon [4]. For General American English, I disabled the flapping rule which neutralizes the contrast between /t/ and /d/ in certain environments, so as to gain more accurate underlying phonological forms. For ease of use, I converted each phoneme to a single, unique symbol. The lexicon contained all POS tags compatible with each entry, plus a full morphological parse. I combined varietal respellings by merging entries which had the same phonological string, morphological structure, and POS.

For frequency information, including the formation of word vectors, I used the Corpus of Contemporary American English, which contains 450 million words of American English [5], and the British National Corpus, which contains 100 million words of British English [6]. I used the tokenization given in each of these corpora to extract bigrams. All tokens were lowercased and I attempted to exclude text describing speakers and section headings by use of heuristics (in COCA, by removing up to 20 words at the start of a line preceding a ¡p¿ tag, which separates headings, and up to 10 words at the start of a sentence preceding a colon (:), which separates speakers; in BNC, by only extracting between certain markup tags).

All of these materials are different from those used by Wedel et al., and are notably larger. However, as I used only 2 of their 9 languages, I had only just over one third the number of mergers, meaning that conclusions about the status of the Functional Load Hypothesis could not be supported; instead, I compared the performance of different weighting schemes.

3.3 Baseline weighting schemes

I used variations on the three weighting schemes employed by Wedel et al. as baselines.

The CONSTANT weighting scheme weights each minimal pair with value 1, as described in Equation (2):

$$W_{con}(o, o') = 1 \tag{2}$$

The FREQUENCY RATIO weighting scheme weights each minimal pair by the ratio of frequency of the the less frequent unigram to that of the more frequent unigram, as described in Equation (3):

$$W_{freq}(o, o') = \frac{\min(c(o), c(o'))}{\max(c(o), c(o'))} \tag{3}$$

The POS AGREEMENT weighting scheme weights each minimal pair by the probability that two random instances of the words have the same POS tag, assuming a uniform distribution over POS tags that are possible for each word, as described in Equation (4):

$$W_{pos}(o, o') = \frac{|POS(o) \cap POS(o')|^2}{|POS(o)||POS(o')|} \tag{4}$$

Note that the POS AGREEMENT weighting scheme is an extension on that employed by Wedel et al.; they had only one POS tag for each word, and weighted a minimal pair by 1 if and only if the two words concerned had the same POS.

3.4 Word vector weighting

I calculated simple word vectors from bigram counts, and used the cosine similarity of these vectors for weighting. I used bigram counts so that the vectors would encode mostly syntactic information, under the assumption that the vector approach should aim to capture (and improve upon) the information encoded in POS tags.

To form the vectors for words in the lexicon, I first identified the top $k = 5000$ words in the corpus by unigram frequency (treating all proper nouns as instances of a single word *PROPER*, as different proper nouns are likely to be sparsely distributed and syntactically uninformative). From each of these context words, I formed two features for the vectors: one indicating that the context word occurred directly to the left of the lexicon word, and one indicating that the context word occurred directly to the right of the lexicon word. Position information was retained in this way so as to promote the syntactic nature of the vectors. I filled in these features with the counts of the (context-word, lexicon-word) bigrams and then converted them to normalized positive pointwise mutual information values following Equation (5), which measure the extent to which two words are collocated above what is expected by chance [7]:

$$I(w_1, w_2) = \left[\ln \frac{\hat{p}(w_1, w_2)}{\hat{p}(w_1)\hat{p}(w_2)} \cdot \frac{-1}{\ln \hat{p}(w_1, w_2)}\right]^+ \tag{5}$$

where $[\cdot]^+ = \max(\cdot, 0)$.

I estimated the unigram and bigram probabilities $\hat{p}(w)$ and $\hat{p}(w_1, w_2)$ respectively two different ways: one with smoothing, to minimize the impact of low-frequency events due to sparsity, and one without. The unsmoothed version normalized the relevant counts by the total number of tokens in the corpus, N, as in Equation (6):

$$\hat{p}(w) = \frac{c(w)}{N} \quad , \quad \hat{p}(w_1, w_2) = \frac{c(w_1, w_2)}{N} \tag{6}$$

The smoothed version employed simple (+1) Laplace smoothing, as in equations (7) (for unigram context words), (8) (for unigram lexicon words) and (9) (for bigram (context-word, lexicon-word) pairs), where $|\mathcal{L}_X|$ represents the number of word types in the lexicon:

$$\hat{p}_{ctxt}(w) = \frac{c(w) + |\mathcal{L}_X|}{N + k|\mathcal{L}_X|} \tag{7}$$

$$\hat{p}_{lex}(w) = \frac{c(w) + k}{N + k|\mathcal{L}_X|} \tag{8}$$

$$\hat{p}(w_1, w_2) = \frac{c(w_1, w_2) + 1}{N + k|\mathcal{L}_X|} \tag{9}$$

The UNSMOOTHED VECTOR weighting scheme weights each minimal pair by the cosine similarity of the unsmoothed vectors $\vec{u}(o)$ and $\vec{u}(o')$ corresponding to the words in the minimal pair, as described in Equation (10):

$$W_{uvec}(o, o') = \frac{\vec{u}(o) \cdot \vec{u}(o')}{\|\vec{u}(o)\| \, \|\vec{u}(o')\|} \tag{10}$$

Similarly, the SMOOTHED VECTOR weighting scheme weights each minimal pair by the cosine similarity of the smoothed vectors $\vec{s}(o)$ and $\vec{s}(o')$ corresponding to the words in the minimal pair, as described in Equation (11):

$$W_{svec}(o, o') = \frac{\vec{s}(o) \cdot \vec{s}(o')}{\|\vec{s}(o)\| \, \|\vec{s}(o')\|} \tag{11}$$

4 Results

Boxplots of the functional load distributions for merged and unmerged contrasts under each weighting scheme are shown in Figure 1 (at the end of the document). As can be seen, the functional load values for merged contrasts are generally lower under all weighting schemes, as expected under the Functional Load Hypothesis.

To assess the extent to which functional load difference predicts merger under each weighting scheme, I followed Wedel et al. in running mixed-effects logistic regression analyses with the natural log of functional load as the independent variable and a random effect of (language, subsystem) pair. Unlike Wedel et al, I did not include phoneme frequency as an additional predictor in the model, as I was concerned about the prospect of overfitting, given that I had so few mergers in my dataset. Instead, I filtered the analysis to only include those contrasts for which both phonemes occurred in the relevant environment in at least 100 different word types. This also had the effect of ensuring that there was at least one minimal pair for each contrast; since Wedel et al. reported that there was no effect of phoneme frequency when minimal pairs existed, I was not concerned about the analysis being misleading due to the omission of a predictor for phoneme frequency.

For each weighting scheme, I report the coefficient (β) and p-value for the functional load predictor, and the AIC of the model in Table 1. The magnitude of β is not informative as each functional load measurement has a different range of appropriate values; however, the sign is informative. The Functional Load Hypothesis predicts that the functional load of merged contrasts will be lower than that of unmerged contrasts, giving rise to a negative β. The lower the p-value, the more confidently the model can be said to support the Functional Load Hypothesis. If β is negative and p is low (e.g. < 0.1, higher than the conventional 0.05 due to the low power associated with the reduced sample size), then AIC values can be used to compare the efficacy of the weighting schemes with respect to the Functional Load Hypothesis. The lower the AIC, the better the quality of the model fit, and thus the more effective the weighting scheme at capturing functional load in a way that supports the Functional Load Hypothesis.

As can be seen, the results under all weighting schemes except the SMOOTHED VECTOR scheme

Table 1: Results of mixed-effects logistic regression models.

Weighting Scheme	β	p	AIC
CONSTANT	-0.311	0.045	128.03
FREQRATIO	-0.297	0.033	127.52
POS	-0.316	0.044	127.57
VECUNSMOOTHED	-0.275	0.066	128.64
VECSMOOTHED	-0.150	0.230	126.61

seem to support the Functional Load Hypothesis. Among these, the FREQUENCY RATIO and POS AGREEMENT schemes seem to be the most effective, followed by the CONSTANT scheme, and then the UNSMOOTHED VECTOR scheme. The difference in efficacy between the UNSMOOTHED VECTOR scheme and the CONSTANT scheme is approximately equal to that between the CONSTANT scheme and the FREQUENCY RATIO and POS AGREEMENT schemes, and each of these differences is small. Though the SMOOTHED VECTOR weighting scheme does not show strong evidence of supporting the Functional Load Hypothesis, it does yield the lowest AIC score. The reason for this is unclear, but note that lower functional load values are assigned under this scheme than under any other; the inability to observe large differences between merged and unmerged contrasts may be related to this more compressed range of scores.

5 Discussion

The results of this experiment replicated those of Wedel et al., despite using fewer languages. Minimal pair count was found to be correlated with the probability of phoneme merger, and weighting the count according to unigram frequencies or POS appeared to yield improvements in the strength of this relationship. Interestingly, raw frequency ratios seemed most effective as weighting scheme here, whereas in Wedel et al.'s study they were only effective when binned, and then less so than POS-based weighting. This may be a result of the different lexica or corpora used here, or may be a quirk from English that is not observed in the other 7 languages that Wedel et al. investigated. It would be valuable to include more languages in the present dataset to investigate how well these results generalize.

The results from word vector weightings were less impressive, though not entirely disheartening. Word vector weighting did not appear to improve upon the constant-weighted minimal pair count in predicting merger probability. Part of the problem may be that the word vectors were high-dimensional (10,000 dimensions) and sparse, causing their similarities to be very small. This could be addressed by using a smaller context set, but that could also lead to a loss of information. An alternative would be to train compressed word vectors using a neural network.

The word vectors may also have suffered from being too restricted. The features in the word vectors were all composed from bigrams, which encode mostly syntactic material. But the Functional Load Hypothesis doesn't rely upon syntactic similarity in any way; it simply states that less two words in a minimal pair are able to be disambiguated from non-phonological factors, the higher the load on the phonological contrast between those words. The motivation for word vectors was that they could identify the extent to which two words are substitutable, i.e. appropriate in the same context, but this appropriateness is governed by semantic factors as well as syntactic ones. So extending the word vectors to include a degree of semantic information, by extending the feature window from 1 word either side to 2, may improve performance.

Finally, it is worth noting that Wedel et al. reported improved results when using a lexicon composed of lemmas (word stems) rather than lexemes (word surface forms). This is something that could be attempted here. Since the English lexicon comes with morphological parses and POS tags, it would be relatively easy to identify words with the same lemmas: they would all have the same prefixes and same root, and may have different suffixes but would all have the same POS tag (as realizations of lemmas differ in terms of inflection only, which doesn't change POS and which is found in suffixes in English). In the general case, however, where no extra information beyond a phonologically-annotated lexicon and a corpus is assumed, this would be very difficult.

6 Conclusion

In this project, I investigated different ways of measuring functional load. I extended measurements used by Wedel et al. on a smaller but richer dataset and replicated their main findings, namely that the Functional Load Hypothesis appears to be quantitatively supported. I then investigated a new measurement utilizing word vectors, which was theoretically and practically motivated, but which did not perform as well as I had hoped.

However, the system that I have built for performing this analysis is extremely general and can easily be extended to incorporate new data and new measurements. Future work can look at including data from other languages to test the generalization of the present results, using lemma-based rather than lexeme-based lexica, using more compressed word vectors, or extending the word vector window to incorporate semantic as well as syntactic information.

References

[1] André Martinet. Function, Structure, and Sound Change. *Word*, 8(1):1–32, 1952.

[2] Andrew Wedel, Abby Kaplan, and Scott Jackson. High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2):179–186, 2013.

[3] Andrew Wedel, Scott Jackson, and Abby Kaplan. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*, 56:395–417, 2013.

[4] Susan Fitt. Unisyn Lexicon, version 1.3, n.d. Available online at http://www.cstr.ed.ac.uk/projects/unisyn/.

[5] Mark Davies. The Corpus of Contemporary American English, 2008. Available online at http://corpus.byu.edu/coca/.

[6] The British National Corpus, version 2 (BNC World), 2001. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/.

[7] Bouma, Gerlof. Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of GSCL*, pages 31–40, 2009.

Table 2: Merged and unmerged phoneme pairs included in the dataset.

Language	Subsystem	Merged pairs	Unmerged pairs
BrE	V ∼ V	aɪ ∼ ɔɪ, ʊə ∼ ɔ, ɪə ∼ ɛə, ɜ ∼ ɛə	i ∼ ɪ, u ∼ ʊ, ʌ ∼ ʊ, ɪ ∼ ɛ, i ∼ ɛ, ɛ ∼ æ, æ ∼ a, u ∼ ɒ, ɒ ∼ ɔ, ɔ ∼ a, ɛ ∼ ɜ, ɜ ∼ ʌ, ʌ ∼ ɔ, eɪ ∼ ɛ, eɪ ∼ ɪ, aɪ ∼ a, aɪ ∼ ɪ, ɔɪ ∼ ɔ, ɔɪ ∼ ɪ, oʊ ∼ ʊ, aʊ ∼ a, aʊ ∼ ʊ, ɪə ∼ ɪ, ɛə ∼ ɛ, ʊə ∼ ʊ, eɪ ∼ aɪ, eɪ ∼ ɔɪ, oʊ ∼ aʊ, aɪ ∼ aʊ, ɛə ∼ ʊə, ɪə ∼ ʊə, eɪ ∼ ɛə
BrE	C ∼ C	θ ∼ t, θ ∼ f, θ ∼ s, ð ∼ d, ð ∼ v, ð ∼ z	p ∼ b, f ∼ v, θ ∼ ð, t ∼ d, s ∼ z, tʃ ∼ dʒ, ʃ ∼ ʒ, k ∼ g, m ∼ n, m ∼ ŋ, n ∼ ŋ, p ∼ t, p ∼ k, t ∼ k, b ∼ d, b ∼ g, d ∼ g, f ∼ s, f ∼ ʃ, f ∼ h, θ ∼ h, s ∼ ʃ, s ∼ h, ʃ ∼ h, v ∼ z, v ∼ ʒ, ð ∼ ʒ, z ∼ ʒ, w ∼ j, l ∼ ɹ, p ∼ f, m ∼ b, m ∼ v, b ∼ v, t ∼ s, t ∼ ʃ, t ∼ tʃ, θ ∼ s, θ ∼ ʃ, θ ∼ tʃ, s ∼ ʃ, s ∼ tʃ, ʃ ∼ tʃ, n ∼ d, n ∼ ð, n ∼ z, n ∼ ʒ, n ∼ dʒ, n ∼ l, n ∼ ɹ, d ∼ z, d ∼ ʒ, d ∼ dʒ, d ∼ l, d ∼ ɹ, ð ∼ z, ð ∼ ʒ, ð ∼ dʒ, ð ∼ l, ð ∼ ɹ, z ∼ ʒ, z ∼ dʒ, z ∼ l, z ∼ ɹ, ʒ ∼ dʒ, ʒ ∼ l, ʒ ∼ ɹ, dʒ ∼ l, dʒ ∼ ɹ, ŋ ∼ g, k ∼ h, g ∼ ʒ, g ∼ z, k ∼ ʃ, k ∼ s, m̩ ∼ n̩, n̩ ∼ l̩
AmE	V ∼ V	ɒ ∼ ɔ, ɜɹ ∼ ɔɪ, ɔ ∼ a	i ∼ ɪ, u ∼ ʊ, ʌ ∼ ʊ, ɪ ∼ ɛ, i ∼ e, i ∼ ɛ, e ∼ ɛ, ɛ ∼ æ, æ ∼ a, u ∼ ɒ, ɛ ∼ ɜ, ʌ ∼ ɔ, aɪ ∼ a, aɪ ∼ ɪ, ɔɪ ∼ ɔ, ɔɪ ∼ ɪ, oʊ ∼ ʊ, aʊ ∼ a, aʊ ∼ ʊ, aɪ ∼ ɔɪ, oʊ ∼ aʊ, aɪ ∼ aʊ
AmE	V ∼ V / _ n	ɪ ∼ ɛ	i ∼ ɪ, u ∼ ʊ, ʌ ∼ ʊ, i ∼ e, i ∼ ɛ, e ∼ ɛ, ɛ ∼ æ, æ ∼ a, u ∼ ɒ, ɒ ∼ ɔ, ɔ ∼ a, ɛ ∼ ɜ, ʌ ∼ ɔ, aɪ ∼ a, aɪ ∼ ɪ, ɔɪ ∼ ɔ, ɔɪ ∼ ɪ, oʊ ∼ ʊ, aʊ ∼ a, aʊ ∼ ʊ, aɪ ∼ ɔɪ, oʊ ∼ aʊ, aɪ ∼ aʊ
AmE	V ∼ V / _ l	i ∼ ɪ, u ∼ ʊ, oʊ ∼ ʊ, ʌ ∼ ɔ, ʌ ∼ ʊ	ɪ ∼ ɛ, i ∼ e, i ∼ ɛ, e ∼ ɛ, ɛ ∼ æ, æ ∼ a, u ∼ ɒ, ɒ ∼ ɔ, ɔ ∼ a, ɛ ∼ ɜ, aɪ ∼ a, aɪ ∼ ɪ, ɔɪ ∼ ɔ, ɔɪ ∼ ɪ, aʊ ∼ a, aʊ ∼ ʊ, aɪ ∼ ɔɪ, oʊ ∼ aʊ, aɪ ∼ aʊ
AmE	C ∼ C	w ∼ ʍ	p ∼ b, f ∼ v, θ ∼ ð, t ∼ d, s ∼ z, tʃ ∼ dʒ, ʃ ∼ ʒ, k ∼ g, m ∼ n, m ∼ ŋ, n ∼ ŋ, p ∼ t, p ∼ k, t ∼ k, b ∼ d, b ∼ g, d ∼ g, ʍ ∼ f, ʍ ∼ θ, ʍ ∼ s, ʍ ∼ ʃ, ʍ ∼ h, f ∼ θ, f ∼ s, f ∼ ʃ, f ∼ h, θ ∼ s, θ ∼ h, s ∼ ʃ, s ∼ h, ʃ ∼ h, v ∼ ð, v ∼ z, v ∼ ʒ, ð ∼ z, ð ∼ ʒ, z ∼ ʒ, w ∼ j, l ∼ ɹ, p ∼ f, p ∼ ʍ, m ∼ b, m ∼ v, b ∼ v, t ∼ θ, t ∼ s, t ∼ ʃ, t ∼ tʃ, θ ∼ s, θ ∼ ʃ, θ ∼ tʃ, s ∼ ʃ, s ∼ tʃ, ʃ ∼ tʃ, n ∼ d, n ∼ ð, n ∼ z, n ∼ ʒ, n ∼ dʒ, n ∼ l, n ∼ ɹ, d ∼ ð, d ∼ z, d ∼ ʒ, d ∼ dʒ, d ∼ l, d ∼ ɹ, ð ∼ z, ð ∼ ʒ, ð ∼ dʒ, ð ∼ l, ð ∼ ɹ, z ∼ ʒ, z ∼ dʒ, z ∼ l, z ∼ ɹ, ʒ ∼ dʒ, ʒ ∼ l, ʒ ∼ ɹ, dʒ ∼ l, dʒ ∼ ɹ, ŋ ∼ g, k ∼ h, g ∼ ʒ, g ∼ z, k ∼ ʃ, k ∼ s, m̩ ∼ n̩, n̩ ∼ l̩

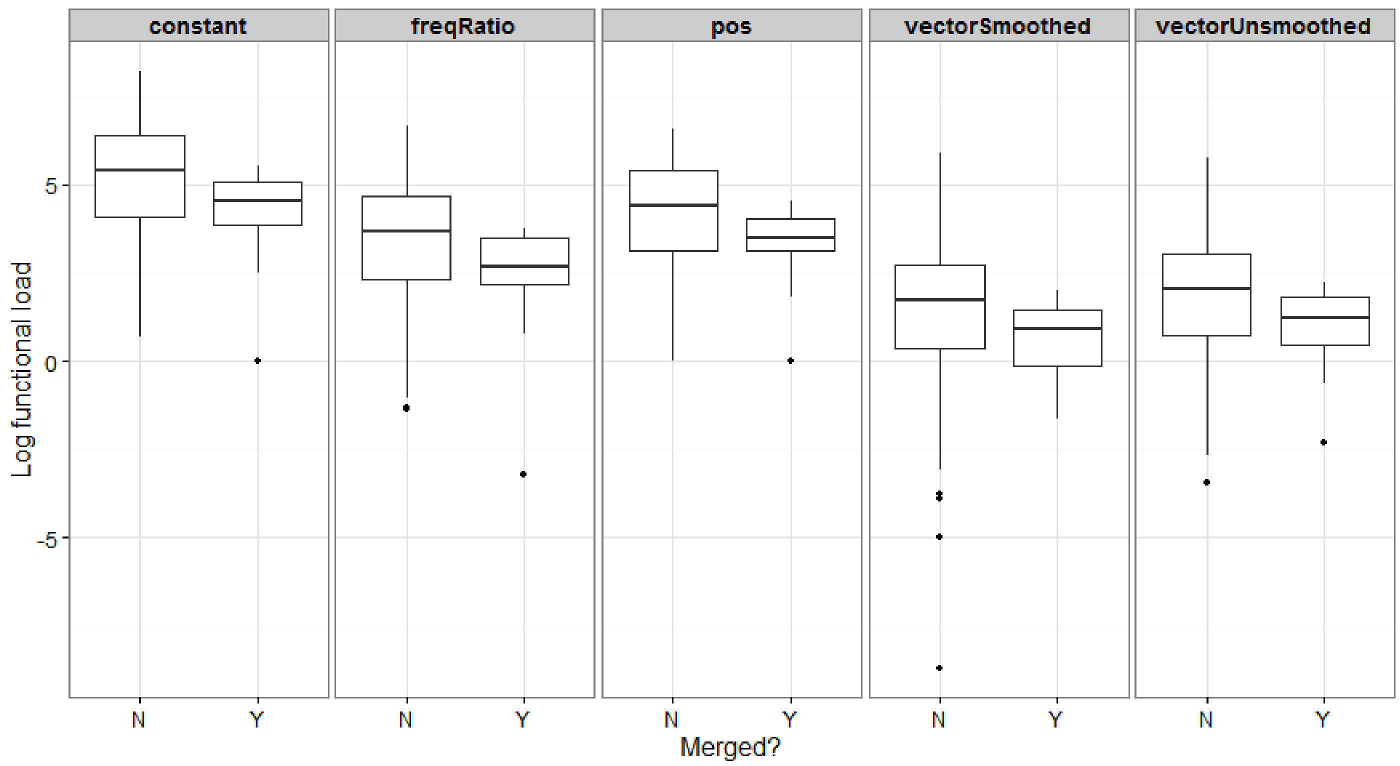

Figure 1: Functional load distributions under different weighting schemes