Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals

Dinoj Surendran and **Partha Niyogi**
Computer Science Department, University of Chicago.

This is a LaTeXed version of the chapter in the upcoming book in: Nedergaard Thomsen, Ole (ed. fc.) Competing Models of Linguistic Change: Evolution and Beyond. In commemoration of Eugenio Coseriu (1921-2002). *Amsterdam & Philadelphia: Benjamins.*

Introduction

Languages convey information using several methods, and rely to different extents on different methods. The amount of reliance of a language on a method is termed the 'functional load' of the method in the language. The term goes back to early Prague School days (Mathesius, 1929; Jakobson, 1931; Trubetzkoy, 1939), though then it was usually taken to refer only to the importance of phonemic contrasts, particularly binary oppositions.

We recently described a general framework to find the functional load (FL) of phonemic oppositions, distinctive features, suprasegmentals, and other phonological contrasts (Surendran and Niyogi, 2003). It is a generalization of previous work on quantifying functional load in linguistics (Hockett, 1955; Wang, 1967) and automatic speech recognition (Carter, 1987).

While still an approximation, it has already produced results not obtainable with previous definitions of functional load. For instance, Surendran and Levow (2004) found that the functional load of tone in Mandarin is as high as that of vowels. This means it is at least as important to identify the tone of a Mandarin syllable as it is to identify its vowels.

King (1967b) notes that Mathesius (1931) "regarded functional load as one part of a complete phonological description of a language along with the roster of phonemes, phonemic variants, distinctive features, and the rest." We agree with this view. While we have an interest in any role functional load might have in sound change, our primary concern here is that a historical linguist who wants to investigate such a role has the computational tools to do so.

The outline of this article is as follows. First, in Section 1, we give an example of how functional load values can be used to investigate a hypothesis regarding sound change. Then, in Sections 2 and 3, we describe a framework for functional load in increasing levels of generality, starting with the limited form proposed by Hockett (1955). Several examples, abstract and empirical, are provided.

1 Example: Testing the Martinet Hypothesis in a Cantonese merger

One factor determining whether phonemes x and y merge in a language is the perceptual distance between them. Another factor, suggested by Martinet (1933), also see Peeters (1992), is the func-

l	p^h	t^h	k^h	p	t	k	w	ts	ts^h	m	h	f	s	ng	k^{wh}	k^w	j	
9	3	1	3	0	1	7	0	3	5	9	3	2	2	1	0	0	4	n
	6	4	3	7	37	70	37	38	24	44	14	8	9	3	0	5	14	l
		1	1	2	3	5	2	2	2	6	2	1	1	1	0	0	2	p^h
			0	2	4	8	2	6	3	1	15	2	24	1	0	2	4	t^h
				1	1	4	0	1	1	0	1	0	1	0	0	0	3	k^h
					7	25	10	7	4	6	9	2	19	1	0	3	39	p
						13	7	31	11	5	36	5	15	1	0	1	7	t
							83	27	12	17	17	8	20	2	0	11	33	k
								8	11	9	15	3	6	1	0	12	5	w
									21	11	17	4	22	1	0	2	25	ts
										15	5	2	10	10	0	2	7	ts^h
											29	3	3	1	0	1	9	m
												2	24	1	0	2	4	h
													7	1	0	1	1	f
														4	0	3	12	s
															0	0	1	ng
																0	1	k^{wh}
																	1	k^w

Table 1: The functional load of all binary consonantal oppositions (in word-initial position) in Cantonese, as computed with a word unigram model on the CANCORP corpus (Lee et al., 1996). All values shown should be multiplied by 0.0001.

tional load $FL(x, y)$ of the x-y opposition i.e. how much the language relies on telling apart x and y. Martinet hypothesized that a high $FL(x, y)$ leads to a lower likelihood of a merger.

The only computational investigation of this hypothesis thus far is that of King (1967a), who found no evidence that it was true. Doubts have been raised to his methodology (Hockett, 1967), and to his overly harsh conclusion that the hypothesis was false. However, while King's work had limitations, it was done in a time of limited computing resources and was a major advance on talking about functional load qualitatively. Sadly, it was not followed up.

A full test of the Martinet Hypothesis requires examples of mergers in different languages, with appropriate (pre-merger) corpora for each case. We only have one example, but this suffices for illustrative purposes.

In the second half of the 20th century, n merged with l in Cantonese in word-initial position (Zee, 1999). For such a recent merger, corpus data is available. We used a word-frequency list derived from CANCORP (Lee et al., 1996), a corpus of Cantonese adult-child speech which has coded n and l as they would have occurred before the merger. It is not a large corpus, and its nature means that there is a higher percentage of shorter words than is normal. However, it is appropriate since mergers are most likely to occur as children learn a language.

Leaving definitions for later, we obtained the value 0.00090 for FL(n,l), where the n-l opposition was only lost in word-initial position. Such a small number might tempt one to conclude that this is indeed an example of the loss of a contrast with low functional load. However, that would be premature, as the absolute value for the load of a contrast is meaningless by itself. It can only become meaningful when compared to loads of other contrasts.

Table 1 shows the FL values for all binary consonantal oppositions in Cantonese, when the opposition was lost only in word-initial position. This gives a much better sense of how small or large FL(n,l) is. However, 'much better' does not mean 'definite', and linguistic knowledge is required to interpret the data. The key question is which of the 171 oppositions of Table 1 should be compared to the n-l opposition. Consider the following possibilities:

1. All 171 oppositions are comparable. Of these, 121 (74%) have a lower FL than the n-l opposition. Thus, the n-l opposition had a moderately high importance compared with consonantal oppositions.

2. On the other hand, several of those pairs seem irrelevant for the purpose of mergers. Perhaps only those pairs that are likely to merge should be considered. While it is not clear what 'likely to merge' means, let us suppose for argument's sake that only consonants that have a place of articulation in common (consonants with secondary articulations have two places) can merge.

   In this case, only coronal consonants should be considered, namely n, l, t, t^h, s, ts, ts^h. Of the 21 binary coronal oppositions, 10 have a larger functional load than the n-l opposition and 10 have a smaller functional load. Thus, the n-l opposition was of average importance compared to other coronal oppositions.

3. Yet a third point to bear in mind for interpretative purposes is that the phoneme that vanished in the n-l merger was n. Resorting to blatant anthropomorphism for a moment, if n had to disappear (in word-initial position), why did it have to merge with l rather than with some other consonant?

   In this case, only consider the 18 oppositions of the form n-x, where x is any consonant other than n. Of these, only FL(n,m) = 0.00091 is higher than FL(n,l). Even when allowing for random variation in the FL values obtained, it is clear that the n-l opposition was very important compared to binary oppositions involving n and other consonants.

There are, of course, other possible interpretations. The key point to note is that functional load values should be interpreted with respect to other functional load values, and the choice of 'other' makes a difference. The most conservative conclusion based on the above observations is that this is an example of the loss of a binary opposition with non-low functional load.

More examples in other languages must be analyzed before we can make further generalizations. We hope we have at least whetted the reader's appetite for functional load data.

2 Defining the functional load of binary oppositions

Binary oppositions of phonemes are the most intuitive kind of phonological contrast. As Meyerstein (1970) noted in his survey of functional load, this was the only type of contrast most linguists attempted to quantify.

Perhaps the most common definition of $FL(x, y)$, the functional load of the x-y opposition, is the number of minimal word pairs that are distinguished solely by the opposition. The major flaw with this definition is that it ignores word frequency. Besides, it is not generalizable to a form that takes into account syllable and word structure or suprasegmentals. We shall say no more about it.

2.1 Hockett's definition

The first definition of $FL(x, y)$ that took word frequency into account was that of Hockett (1955). He did not actually perform any computations with this definition, although Wang (1967) did.

The definition was based on the information theoretic methods introduced by Shannon (1951), and assumes that language is a sequence of phonemes whose entropy can be computed. This sequence is infinite, representing all possible utterances in the language. We can associate with a language/sequence L a positive real number $H(L)$ representing how much information L transmits.

Suppose x and y are phonemes in L. If they cannot be distinguished, then each occurrence of x or y in L can be replaced by one of a new (archi)phoneme to get a new language L_{xy}. Then the functional load of the x-y opposition is

$$FL(x, y) = \frac{H(L) - H(L_{xy})}{H(L)} \tag{1}$$

This can be interpreted as the fraction of information lost by L when the x-y opposition is lost.

2.2 Computational Details

It is not possible to use (1) in practice. We now give the details of how it can be made usable, taking care to note the additional parameters that are required.

To find the entropy $H(L)$ of language/sequence L, we have to assume that L is generated by a stationary and ergodic stochastic process (Cover and Thomas, 1991). This assumption is not true, but is true enough for our purposes. We need it because the entropy of a sequence is a meaningless concept — one can only compute the entropy of a stationary and ergodic stochastic process. Therefore, we define $H(L)$ to be the entropy of this process or, more precisely, the entropy of the process's stationary distribution.

Intuitively, this can be thought of as follows: suppose there are two native speakers of L in a room.

When one speaks, i.e. produces a sequence of phonemes, the other one listens. Suppose the listener fails to understand a phoneme and has to guess its identity based on her knowledge of L. $H(L)$ refers to the uncertainty in guessing; the higher it is, the harder it is to guess the phoneme and the less redundant L is.

Unfortunately, we will never have access to all possible utterances in L, only a finite subset of them. This means we must make more assumptions; that L is generated by a k-order Markov process, for some finite non-negative integer k. This means that the probability distribution on any phoneme of L depends on the k phonemes that occurred before it.

In our speaker-listener analog above, this means that the only knowledge of L that the listener can use to guess the identity of a phoneme is the identity of the k phonemes preceding it and the distribution of $(k+1)$-grams in L. An n-gram simply refers to a sequence of n units, in this case phonemes. The uncertainty in guessing, with this limitation, is denoted by $H_k(L)$, and decreases as k increases. A classic theorem of Shannon (1951) shows that $H_k(L)$ approaches $H(L)$ as k becomes infinite.

The finite subset of L that we have access to is called a corpus, S. This is a large, finite sequence of phonemes. As S could be any subset of L, we have to speak of $H_{kS}(L)$ instead of $H_k(L)$. If X_{k+1} is the set of all possible $(k+1)$-grams and D_{k+1} is the probability distribution on X_{k+1}, so that each $(k+1)$-gram x in X has probability $p(x)$, then

$$H_{kS}(L) = \frac{1}{k+1}\left(-\sum_{x\in X} p(x)\log_2 p(x)\right) \tag{2}$$

There are several ways of estimating D_{k+1} from S. The simplest is based on unsmoothed counts of $(k+1)$-grams in S. Suppose $c(x)$ is the number of times that $(k+1)$-gram x appears in S, and $c(X_{k+1}) = \sum_{x\in X_{k+1}} c(x)$. Then

$$p(x) = \frac{c(x)}{c(X_{k+1})} \tag{3}$$

To illustrate, suppose we have a toy language K with phonemes a, u and t. All we know about K is in a corpus S = "atuattatuatatautuaattuua". If we assume K is generated by a 1-order Markov process, then $X_2 = \{$ aa, at, au, ta, tt, tu, ua, ut, uu $\}$ and $c(\text{aa}) = 1$, $c(\text{at}) = 6$, $c(\text{au}) = 1$, $c(\text{ta}) = 3$, $c(\text{tt}) = 2$, $c(\text{tu}) = 4$, $c(\text{ua}) = 4$, $c(\text{ut}) = 1$, $c(\text{uu}) = 1$. The sum of these counts is $c(X_2) = 23$. D_2 is estimated from these counts: $p(aa) = \frac{1}{23}$, $p(at) = \frac{6}{23}$, etc. Finally $H_{1,S} = \frac{1}{2}[\frac{1}{23}\log_2\frac{23}{1} + \frac{6}{23}\log_2\frac{23}{6} + \ldots + \frac{1}{23}\log_2\frac{23}{1}] = \frac{1}{2}(2.86) = 1.43$.

In other words, a computationally feasible version of (1) is :

$$FL_{kS}(x,y) = \frac{H_{kS}(L) - H_{kS.xy}(L_{xy})}{H_{kS}(L)} \tag{4}$$

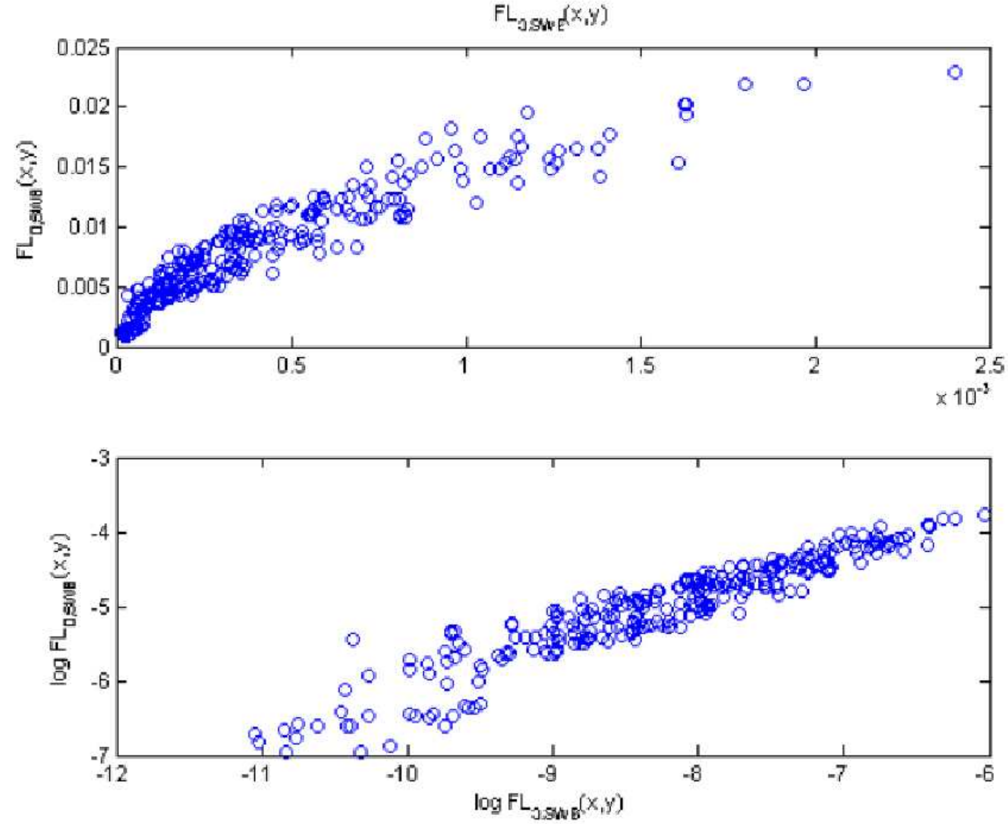

Figure 1: Comparing functional load values, and their logarithms, obtained with k=0 and k=3 on the vertical and horizontal axes respectively, of American English from the ICSI subset of the Switchboard Corpus (Greenberg, 1996). The correlation of the FL values is 0.928 ($p << 0.001$) and of the log FL values is 0.945 ($p << 0.001$).

$S.xy$ is the corpus S with each occurrence of x or y replaced by that of a new phoneme. It represents L_{xy} in the same way that S represents L. $FL_{kS}(x, y)$ can no longer be interpreted as the fraction of information lost when the $x - y$ opposition is lost, as such an interpretation would only be true if L was generated by a k-order Markov process. However, by comparing several values obtained with the same parameters, as we did with the Cantonese merger example of the previous section, we can interpret this value relatively.

Returning to our toy example, suppose we want to know the functional load of the a-u opposition with the same k and S. We create a new corpus $S.\mathrm{au}$ with each a or u replaced by a new phoneme V. Then $S.\mathrm{au}$ = "VtVVttVtVVtVtVVtVVVttVVV", $c(\mathrm{Vt}) = 7$, $c(\mathrm{VV}) = 7$, $c(\mathrm{tt}) = 2$, $c(\mathrm{tV}) = 7$, and eventually $H_{1,S.\mathrm{au}} = \frac{1}{2}[\frac{7}{23}\log_2\frac{23}{7} + \frac{7}{23}\log_2\frac{23}{7} + \frac{2}{23}\log_2\frac{23}{2} + \frac{7}{23}\log_2\frac{23}{7}] = \frac{1}{2}(1.87) = 0.94$. Then the functional load $FL_{1,S}(\mathrm{a}, \mathrm{u}) = (1.43 - 0.94)/1.43 = 0.34$.

2.3 Robustness to k (Markov order)

It would be nice to have some assurance that the values used for k and S in (4) make little difference to our interpretation of the values we get. Surprisingly, there has been no mention, let alone study, of this problem in the functional load literature. This may be because it is mathematically clear that different choices of k and S (e.g. different k for the same S) result in different FL values.

However, there is a loophole. We have already said that FL values should be interpreted relative to other FL values. Once we accept this relativity, then preliminary experiments suggest that interpretations are often robust to different choices of k and S.

For example, we computed the functional load of all consonantal oppositions in English with $k = 0$ and $k = 3$ using the ICSI subset of the Switchboard corpus (Godfrey et al., 1992; Greenberg, 1996) of hand-transcribed spontaneous telephone conversations of US English speakers. Figure 1 shows how $FL_{0,Swbd}(x, y)$ and $FL_{3,Swbd}(x, y)$ compare for all pairs of consonants x and y. The correlation is above 0.9 ($p << 0.001$), indicating that one is quite predictable from the other. This is surprising, since the $k = 0$ model does not use any context at all, and is simply based on phoneme frequencies.

2.4 Generalizing to sequences of units other than phonemes

The problems with modeling language as a sequence of phonemes are manifold. There is no way to account for prosody, tones, syllable structure, word structure, phoneme deletion/insertion, etc.

Many of these problems can be fixed by modeling language as a sequence of discrete units of some type, such as phonemes, morphemes, syllables, or words. We shall call an arbitrary type T, and a unit of that type a T-unit. Much sophistication can go into the definition of a type: for example, a word can have several components representing its phonemic and prosodic (and even syntactic and semantic) structure.

This permits the kind of hierarchical definition advocated by Rischel (1961) and and implemented to a limited extent by Kucera (1963). It also permits us to find the functional load of a much larger class of phonological contrasts than previously envisaged. It does not get around the problems of cohort-based language variability models pointed out by Wittgott and Chen (1993).

Everything said in the definition above for phonemes can be said for T-units. This means that there are now three parameters going into the definition of H and FL, and we must speak of $H_{TkS}(L)$ and $FL_{TkS}(x, y)$ instead of $H(L)$ and $FL(x, y)$. The formula (4) is now

$$FL_{TkS}(x, y) = \frac{H_{TkS}(L) - H_{TkS.xy}(L_{xy})}{H_{TkS}(L)} \tag{5}$$

Table 2 shows the functional load of all binary consonantal oppositions in American English using the Switchboard corpus, with T = 'syllable' and $k = 0$. A syllable here is just a phoneme sequence.

p	b	t	d	k	g	C	J	f	v	T	D	s	z	S	Z	m	n	N	l	r	j	h	w	
2	3	6	4	4	2	1	1	3	3	1	3	6	4	2	0	5	7	1	4	4	3	2	5	x
	4	5	4	4	2	2	1	4	4	2	4	5	3	2	1	5	5	0	4	3	3	3	5	p
		5	5	5	3	2	1	3	3	2	6	6	3	2	1	5	7	0	5	5	4	3	7	b
			9	7	4	3	2	5	5	2	6	10	6	3	1	8	12	2	8	7	6	4	9	t
				7	3	2	2	5	5	2	6	9	6	3	1	6	9	1	7	6	5	4	8	d
					4	2	2	5	5	2	6	8	5	2	0	3	5	0	3	3	3	2	4	k
						1	1	2	2	1	3	5	3	2	0	3	5	0	3	3	3	2	4	g
							1	2	2	1	2	2	2	1	0	2	2	0	2	2	2	1	2	C
								1	1	1	2	2	2	1	1	2	2	0	2	1	2	1	2	J
									4	1	3	5	4	2	1	5	6	1	4	4	3	2	5	f
										2	4	6	5	2	1	5	7	1	5	4	3	2	6	v
											2	2	2	1	0	2	2	0	2	2	1	1	2	T
												7	4	2	1	6	8	0	5	5	5	4	8	D
													8	3	1	9	13	2	8	7	5	4	9	s
														2	1	5	10	2	6	5	3	2	6	z
															1	3	3	0	3	2	2	2	3	S
																1	1	0	1	1	1	0	1	Z
																	11	1	7	7	4	4	8	m
																		3	12	10	6	5	11	n
																			1	1	0	0	1	N
																				9	4	3	9	l
																					4	3	8	r
																						4	5	j
																							4	h

Table 2: $FL(x, y)$ for all pairs x, y of consonants in American English, based on syllable unigram data from the ICSI subset of the Switchboard corpus. C and J are the un/voiced affricate 'ch' and 'dzh' respectively, T and D the un/voiced alveolar fricatives 'th' and 'dh', S and Z the un/voiced alveolar sibilants 'sh' and 'zh', N the velar nasal 'ng' and x the glottal stop. All values should be multiplied by 0.001.

2.5 Robustness to corpus used

It is a plain fact that the entropy of a language depends on the corpus used - it can even be used to distinguish between authors in the same language (Kontoyannis, 1993). However, as functional load values are a ratio of entropies, and are to be interpreted relatively anyway, we can hope they will not be as corpus-dependent as raw entropy values.

To test this, we recomputed the values in Table 2 with CELEX (Baayen et al., 1995), a very different source of corpus data. The correlation was 0.797 ($p << 0.001$), which is good, but not entirely satisfactory. However, the agreement is much better for binary oppositions of obstruents, the correlation being 0.892 ($p << 0.001$).

There is an important subtlety hiding here, because syllables in CELEX are different from those in Switchboard. CELEX syllables have two parts instead of one. The first is the phonemic part as before, while the second is a stress part that can have one of the values <primary>, <secondary>

and <unstressed>. Thus the syllables ('pirz',<primary>) would still be distinguishable from ('parz',<unstressed>) when the a-i opposition was lost, but not from ('parz',<primary>).

This means that the 0.797 and 0.892 figures above were computed with the same k and different S and T. To make a comparison with the same k and T but different S, we redid the experiment with the stress values from CELEX ignored. Then the corresponding figures are 0.816 and 0.920 respectively.

This agreement, especially for obstruents, is quite remarkable given the differences between the Switchboard and CELEX corpora. Switchboard has about 36 000 syllable tokens of 4000 types, while CELEX is a word-frequency list derived from a corpus (Birmingham/COBUILD) with 24 000 000 syllable tokens of 11 000 types. Switchboard syllables are based on spontaneous speech of American English, and thus have far fewer consonant clusters than CELEX syllables, which are based on canonical pronunciations of British English. Frequency values for Switchboard are based on spoken language, while those from CELEX are derived mostly from written texts.

This is very good news for historical linguists, as available corpora of historical languages represent written rather than spoken texts, and pronunciations are at best canonical ones.

3 Defining the functional load of general phonological contrasts

Suppose we are computing functional load values with parameters k, S and T, that is, assuming that the language in a corpus S is a sequence of T-units generated by a k-order Markov process. $X = X_1$ is the set of all T-units.

Let $f : X \rightarrow Y$ be any function on X. The range Y of f, can be considered to be a set of units of a new type U. Then the functional load $FL(f)$ of f is defined as :

$$FL_{TkS}(f) = \frac{H_{TkS}(L) - H_{Ukf(S)}(f(L))}{H_{TkS}(L)} \quad (6)$$

The function f represents the loss of the contrast we wish to find the functional load of, and $f(L)$ and $f(S)$ represent the language L and corpus S after the loss of the contrast.

For example, consider the only contrasts we have dealt with so far: the binary opposition of two phonemes p and q. If $T =$ 'phoneme', then X is the set of phonemes, so that $p, q \in X$. We define Y to be X with p and q removed, and a new phoneme p' added. If we define a function $g : X \rightarrow Y$ by $g(p) = g(q) = p'$, and $g(x) = x$ for any x in $X - \{p, q\}$, then $FL_{TkS}(g) = FL_{TkS}(p, q)$ as before.

What if T is not a phoneme? Suppose $T =$ 'syllable', where a syllable $(x_1 \ldots x_b, s)$ has both a phoneme sequence $x_1 \ldots x_b$ and stress component s. We can define a function h on the set of syllables that takes such a syllable to $(g(x_1) \ldots g(x_b), s)$ where g is the function of the previous paragraph. Then $FL_{TkS}(h) = FL_{TkS}(p, q)$.

	Dutch		English		German		Mandarin	
	Syll	Word	Syll	Word	Syll	Word	Syll	Word
Place	67	11	73	20	61	13	65	14
Manner	27	5	39	11	27	8	34	6
Voicing	30	3	23	5	21	1		
Aspiration							17	3
Nasality	15	2	12	3	16	2	8	3i

Table 3: The functional load of some distinctive features in four languages, based on unigram ($k = 0$) counts of syllables and words. Data summarized from Surendran and Niyogi (2003). All values shown should be multiplied by 0.001.

And if T is a word, where a word is a sequence of syllables of the form $s_1 \ldots s_c$ then the required function takes this to $h(s_1) \ldots h(s_c)$. Note that the positive integer c is different for different words.

The generalization to draw here is that any function from phonemes to phonemes induces one from syllables to syllables, which in turn induces one from words to words.

3.1 The functional load of a distinctive feature

Phonemes can be described in terms of distinctive features (Jakobson and Halle, 1956), which do not have to be binary. In the absence of a distinctive feature, certain phonemes would sound alike, and the function f should be defined so that such phonemes are 'collapsed' into a single phoneme.

For example, without aspiration, a Mandarin speaker would be unable to distinguish between ts and ts$^{\mathrm{h}}$, p and p$^{\mathrm{h}}$, t and t$^{\mathrm{h}}$, etc. The functional load of aspiration is defined by $FL(f)$, where f is a function defined by $f(\mathrm{ts}) = f(\mathrm{ts}^{\mathrm{h}}) = \mathrm{ts}'$, $f(\mathrm{p}) = f(\mathrm{p}^{\mathrm{h}}) = \mathrm{p}'$, $f(\mathrm{t}) = f(\mathrm{t}^{\mathrm{h}}) = \mathrm{t}'$, $\ldots$ and $f(x) = x$ if x is any phoneme that is not part of an aspirated-unaspirated pair of consonants.

Another example: without manner, an English speaker would be unable to tell apart b from m from w or d from dh from n, etc. The functional load of place in English is defined by $FL(f)$ where $f(\mathrm{b}) = f(\mathrm{m}) = f(\mathrm{w}) = \mathrm{b}'$, $f(\mathrm{d}) = f(\mathrm{dh}) = f(\mathrm{n}) = \mathrm{d}'$, $f(\mathrm{t}) = f(\mathrm{th}) = \mathrm{t}'$, $\ldots$, and $f(x) = x$ for any other phoneme x.

Functional load values for some distinctive features in Dutch, English, German and Mandarin appear in Table 3. Place is more important than manner in all four languages.

3.2 The functional load of a suprasegmental feature

By suprasegmentals, we refer primarily to stress and, in tonal languages, tone. The definition of these terms is by no means standard.

Return to the situation where the types of units we are dealing with are syllables with both a phoneme sequence and stress component. Define a function f that takes a syllable $(x_1 \dots x_b, s)$ to $(x_1 \dots x_b)$, i.e. ignores the value of its stress component. Then $FL(f)$ is the functional load of stress.

In some tonal languages, syllables can be assumed to have three components: phonemic, stress and tone. Then the functional load of stress is FL(f), where f is a function that takes a syllable $(x_1 \dots x_b, s, t)$ to $(x_1 \dots x_b, t)$ and the functional load of tone is $FL(g)$, where g is a function that takes a syllable $(x_1 \dots x_b, s, t)$ to $(x_1 \dots x_b, s)$.

If words are modeled as sequences of syllables, then f and g above induce word-converting functions whose functional load is what we require.

3.3 The functional load of a condition-dependent contrast

This is best described with a couple of non-trivial examples.

In the Cantonese merger example several pages ago, we computed the functional load of binary phonemic oppositions in word-initial position. No previous definition of functional load suggested how one might deal with conditional loss of contrasts.

We modeled Cantonese as a sequence of words, with each word as a sequence $s_1 \dots s_c$ of syllables, and each syllable $(p_1 \dots p_b, t)$ as a sequence of phonemes with a tone. Note that c and b vary with word and syllable respectively. We then defined f so that it converted a word $s_1 \dots s_c$ to $g(s_1)s_2 \dots s_c$, where g converted a syllable $(x_1 \dots x_b, t)$ to $(h(x_1)x_2 \dots x_b, t)$. When the binary opposition in question was of phonemes p and q, h was defined as $h(p) = h(q) = p'$ and $h(x) = x$ for every other phoneme x.

In other words, the $p - q$ opposition was only lost when the first phoneme in a word was p or q. Recall that g would convert $(x_1 \dots x_b, t)$ to $(h(x_1) \dots h(x_b), t)$ and f would convert a word $s_1 \dots s_c$ to $g(s_1) \dots g(s_c)$ if f was supposed to represent the regular (everywhere) loss of the x-y opposition.

For another example, suppose we represented English as a sequence of syllables and we wanted to represent vowel reduction. This is the loss of distinction between vowels in unstressed syllables. Then we would define f so that it converted a syllable $(x_1 \dots x_b, s)$ to $(g(x_1) \dots g(x_b), s)$ if $s =$ 'unstressed' and to $(x_1 \dots x_b, s)$ otherwise. The function g converts x to x if it is a consonant and to V if it is a vowel. Then $FL(f)$ is the functional load of being able to distinguish between vowels in unstressed syllables.

4 Summary

We have outlined a method for providing quantitative data on how much a language relies on phonemic opposition, distinctive feature, or suprasegmental feature, even when the opposition/feature is lost only in certain conditions. Initial tests suggest it is reasonably robust, even with non-ideal representations of spoken language such as word-frequency lists with canonical pronunciations of words and frequencies from written texts. While it needs to be improved both statistically and linguistically, it can be used in its present state as a tool at the intersection of historical and corpus linguistics.

This article is meant to be a more accessible version of Surendran and Niyogi (2003). The reader seeking more computational details is referred there and to
http://people.cs.uchicago.edu/~dinoj/research/fload .

We would like to thank Stephanie Stokes for pointing us to the Cantonese data and Bert Peeters for useful discussions on how functional load is viewed in the Martinet tradition.

References

R. Harald Baayen, Richard Piepenbrock, and Leon Gulikers. 1995. The CELEX Lexical Database (Release 2) [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia PA.

David Carter. 1987. Information-theoretical Analysis of Phonetic Dictionary Access. *Computer Speech and Language*, 2:1–11.

Thomas M. Cover and Joy A. Thomas. 1991. *Elements of Information Theory.* Wiley-Interscience, New York, NY.

John J. Godfrey, Edward C. Holliman, and Jane McDaniel. 1992. Switchboard: Telephone Speech Corpus for Research and Development. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 517–520.

Steven Greenberg. 1996. The Switchboard Transcription Project: Large Vocabulary Continuous Speech Recognition. Summer Research Workshop Technical Report 24, Johns Hopkins University, Baltimore, MD.

Charles F. Hockett. 1955. *A Manual of Phonology (= International Journal of American Linguistics 21.* Indiana University Press.

Charles F. Hockett. 1967. The Quantification of Functional Load. *Word*, 23:320–339.

Roman Jakobson and Morris Halle. 1956. *Fundamentals of Language.* Mouton, The Hague.

Roman Jakobson. 1931. Prinzipien der Historischen Phonologie. *Travaux du Cercle Linguistique de Prague*, 4:246–267.

Robert King. 1967a. Functional Load and Sound Change. *Language*, 43.

Robert King. 1967b. A Measure for Functional Load. *Studia Linguistica*, 21:1–14.

Ioannis Kontoyannis. 1993. The Complexity and Entropy of Literary Styles. Technical Report 97, Stanford University.

Henry Kucera. 1963. Entropy, Redundancy and Functional Load. In *American Contributions to the Fifth International Conference of Slavists*, pages 191–219, Sofia, Bulgaria. The Hague: Mouton.

Thomas Hun-Tak Lee, Colleen Wong, Samuel Leung, Patrician Man, Alice Cheung, Kitty Szeto, and Cathy S. P. Wong. 1996. The Development of Grammatical Competence in Cantonese-speaking Children : Report of a Project funded by Research Grants Council 1991-4. Technical Report, Chinese University of Hong Kong.

André Martinet. 1933. Remarques sur le Système Phonologique du Francais. *Bulletin de la Société de Linguistique de Paris*, 34:191–202.

Vilem Mathesius. 1929. La Structure Phonologique du Lexique du Tchèque Moderne. *Travaux du Cercle Linguistique de Prague*, 1:67–84.

Vilm Mathesius. 1931. Zum Problem der Belastungs- und kombinationsfähigkeit der Phoneme. *Travaux du Cercle Linguistique de Prague*, 4:148–152.

Rud S. Meyerstein. 1970. Functional Load: Descriptive Limitations, Alternatives of Assessment and Extensions of Application. *Janua Linguarum, Series Minor*, 99.

Bert Peeters. 1992. *Diachronie, Phonologie et Linguistique Fonctionnelle (= Bibliothèque des Cahiers de l'Institut de Linguistique de Louvain, 64).* Peeters, Louvain-la-Neuve.

Jrgen Rischel. 1961. On Functional Load in Phonemics. *Statistical Methods In Linguistics*, 1:1–32.

Claude E. Shannon. 1951. Prediction and Entropy of Printed English. *Bell Systems Technical Journal*, 30:50–64.

Dinoj Surendran and Gina-Anne Levow. 2004. The Functional Load of Tone in Mandarin is as High as that of Vowels. In *Proceedings of the International Conference on Speech Prosody 2004*, pages 99–102, Nara, Japan.

Dinoj Surendran and Partha Niyogi. 2003. Measuring the Usefulness (Functional Load) of Phonological Contrasts. Technical Report TR-2003-12., Department of Computer Science, University of Chicago.

Nikolai Trubetzkoy. 1939. Grundzüge der Phonologie. *Travaux du Cercle Linguistique de Prague*, 7.

William S-Y. Wang. 1967. The Measurement of Functional Load. *Phonetica*, 16:36–54.

Margaret Wittgott and Francine Chen. 1993. *Computational Models of American Speech (= CSLI Lecture Notes, 32).* Center for the Study of Language and Information, Stanford, CA.

Eric Zee. 1999. Change and Variation in the Syllable-initial and Syllable-final Consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics*, 27:120–167.