7-9-2014

Naming Practices in J.R.R. Tolkien's Invented Languages

Wendy Baker et al.
Brigham Young University, wendy_baker@byu.edu

Naming Practices in J.R.R. Tolkien's Invented Languages

Wendy Baker-Smemoe
Brad Wilcox
Bruce Brown
Paul Hoskisson
Sky Rodio Nuttall
Brigham Young University

Introduction

Studies examining personal names usually focus on their etymology or cultural significance (i.e., Lombard 2011; Nummila 2007; Suzman 1994). The requirement that names be "meaningful" characteristic of Zulu, fore example—that is, reflecting "social/financial position in life" (e.g., *uBuyiswanokutthula*, "silence returned"), the circumstances of the child's birth (*uLindiwe,* "awaited"), as well as other matters—is very common across naming practices of the world (Suzman 1994). Other societies use personal names to teach culturally important lessons or remind their owners of important historical events (Lombard 2011). Additional reasons for choosing names, at least historically, include naming the child with hoped for characteristics (e.g., beautiful, handsome) (Rubio Orecilla 2006), or focusing on the location of the birth (Wenzel 2001), or the occupation of the parents (Nummila 2007). Today many names, especially in Western societies, have lost their historical meaning (such as *Cooper* or *Baxter)* although some names have not (such as *Charity* or *Hope*). The significance of personal names can be examined considering their origins and phonotactic properties, in both authentic and fictional settings.

Origin of Names

Of course, with the interaction of many different cultures and peoples, personal names may derive from several sources (Cieslikowa 2001; Lamaj & Memishaj 2001; Leglise & Migge 2006). In fact, many people have at least two names, one chosen from the native language of the parents and the other from the culture where they reside (Aceto 2002). Thus new names constantly become part of the "choosing pool" from which parents decide what to name their child. Names can develop from many origins; a number of names used in Western societies derive from circumstances such as invasions of or trading with other cultures (Verstraten 2006), conversion to Christianity (Stefanov 1991), Pagan mythology (Superanskaya 1999), or outright coinage (Rymes 1996). What causes one particular name to be chosen over another, of course, varies greatly from culture to culture and individual to individual. However, one thing is clear—personal names in most societies today derive from many different languages, motivations, cultures, and people.

Due to these tendencies to vary and combine, the phonological features of personal names can differ significantly (Maduagwu 2010; Nemer 1987). The phonology of personal names may have personal and/or social significance—native English speakers prefer, for example, female names that follow specific features (such as ending in /i/ as in *Jenny* or /n/ as in *Shannon*) and male names that follow others (such as ending in /r/ as in *Connor* or in obstruents as in *Brad*); names will change or fall out of favor if they do not follow these features (such as *Ashley* for a boy's name). Other cultures prefer to keep the exact or very similar phonology to the pronunciation of the name in the borrowed language, even if it violates native language phonotactics (Nemer 1987). Indeed, researchers have

found that personal names often have their own phonology that differs from the common patterns found in other nouns (Hough 2000). Thus, personal names come from a variety of sources, and this variety is also apparent in the phonology of these names.

Phonotactic Properties of Names

Wilcox et al. (2013) studied the degree to which personal names differ in terms of their phonology by examining a corpus of the 100 most prevalent male personal names in English in the nineteenth century using the *phonotactic calculator,* created by Vitevitch and colleagues (Vitevitch & Luce 2004). This calculator determines the relative probability of each phoneme in a word occurring in the location that it does, and also the probability of occurring adjacent to the preceding and following sounds in the word biphones or bi-phonemes as defined by Vitevitch and Luce (2004). In addition, a word's probability is determined by examining its neighborhood density. Words with lots of neighbors (*cap*, *cat*, *can*, *cash*) score higher on the probability calculator than words with few neighbors (*oriole*, *flask*, etc.). The calculator determines neighborhood density and sound location probability by comparing a selected word to calculations already performed on the Francis-Kucera corpus (Kucera & Francis, 1967). For example, in the word *box* (/baks/)*,* the calculator would determine how probable it is that a word would start with the sound /b/, be followed by the sound /a/ in the second position, how likely a word would have /a/ in the second position and /k/ following it and so on. Wilcox et al. (2013) indeed did find that the phonotactic probabilities of each name in the corpus of nineteenth-century century names varied greatly.

Wilcox et al. (2013) also found that personal names can differ from each other in the number and types of phonemes they use (and their frequency compared to other nouns). For example, they found that personal names differ greatly in what phonemes they use, so the frequency of any one phoneme is quite low. Although not examined in their study, personal names could also differ in their syllable structure and phonotactics. Languages can differ greatly in the syllable structures that they allow and what consonants and vowels can occur in what positions in a syllable (Fudge & Shockey 1998). Therefore, personal names borrowed from another language could use only a subset of the possible syllable structures of the language that they were borrowed into. Conversely, the syllable structure of a name may be changed if it violates the syllable structure of the language it is borrowed into (such as changing one's name from *Steven* to *Esteban* when speaking Spanish).

Fictional Names

But what happens when only one source creates many personal names, such as names chosen for a work of fiction? Can one author imitate the phonological diversity found in a corpus of personal names from an actual language? This creation of new names would be particularly likely when authors created fantasy or science fiction—inventing names that supposedly come from a created language or a different world. Fictional authors have been shown to choose names that match the personality of the character or that bring up stereotypes or archetypes (Mhlambi 2007). Black and Wilcox (2011) interviewed several authors and found that other methods of choosing names for fictional characters included choosing names with personal significance for the author, doing research to find names that are unique or significant, and choosing names with phonology that appealed to the authors. Thus choosing names in fiction may have even more social significance than selecting names in general.

Previous research has demonstrated that authors can be identified by examining the syntax or lexicon they use, what in some studies has been called a "wordprint" based on the idea that it is

like a fingerprint for author identification (Grieve 2007; Hilton 1990; Morton 1979). Despite many criticisms of the method and the difficulty of collecting sufficient and adequate data (Croft 1981), this method of examining an author's use of specific syntactic structures, type-token ratios, and other lexical features is often used to identify or verify authorship of documents (Baily 1979; Holmes 1994; Zheng, Quin, Huang, & Chen 2003; Iqbal, Kahn, Fung, & Debbabi 2010). In fact, an author can be identified based on his or her use of function words alone (Miranda Garcia & Calle Lambert 2007). These findings suggest that authors have certain biases toward using specific words and syntactic structures. One linguistic feature that has not been examined sufficiently in author attribution studies is whether authors use some specific sounds (phonemes) more than others: consciously or subconsciously putting sounds together in consistent individual patterns when they invent words. These patterns may create a specific "sound print" (or *phonoprint*).

Tolkien's Character Names

The work of J.R.R. Tolkien offers an excellent test case of whether an author has a phonoprint, or bias towards certain sound sequences, when creating personal names, much as authors demonstrate biases towards specific syntactic structures. A professor of philology and Anglo Saxon at Oxford, Tolkien created several languages based on natural languages and endowed each with a history and culture. It has been said that Tolkien created Middle Earth in order to have a place where his languages "could run free" (Bianco 2004). Thus he purposely attempted to use different sound systems for each of his created languages, and he created names of characters that followed the phonotactics and other patterns of these languages.

In the study that investigated prevalent nineteenth-century English male names, Wilcox et al. (2013) compared the phonological diversity found in nineteenth-century English personal names with that in names created by Tolkien as well as a lesser known author. This study determined that both authors failed to imitate the phonological diversity found in the actual personal names. Each author tended to use the same phonemes repeatedly for different names (as determined by both their phonotactic probability and their frequency counts), thus demonstrating a phonoprint. However, Tolkien had greater diversity in the sounds he used and in the positions in which he placed them in the words than did the lesser-known author. These findings, therefore, seemed to suggest that Tolkien was more adept at creating names following patterns (or lack of them) found in actual naming systems.

One drawback of this study was that all of Tolkien's names were examined as a single group, even though Tolkien created distinct language groups from which he derived these names. It may be that some of his names that were based on actual languages did in fact achieve phonological diversity similar to that found in actual naming systems. Moreover, the previous study did not examine the frequency with which certain phonemes were used with each distinct group of names, nor whether these groups of names differed in terms of syllable structure. Thus the purpose of this current study is to analyze and compare the phonotactic patterns and probabilities (i.e., the phonoprint) of the five different language groups from which Tolkien derived names for his characters (elf, man, dwarf, hobbit, other).

Research Questions

The following research questions guided our investigation:
1. Do the five groups of Tolkien's names differ in terms of overall phoneme probability and overall biphone probability?

2. Do the five groups of names differ in terms of phoneme frequency counts?
3. Do the five groups of names differ in terms of syllable structure?

Tolkien's Languages

Much research has been devoted to Tolkien's languages in general and the personal names derived from these languages in particular. To understand the languages of Tolkien, it is helpful to understand the worlds he created for them to be spoken in. In Tolkien mythology, at the beginning of time were the Valar, from whom two distinct groups of Elves descended, the Western and the Eastern Elves. This distinction is important since Tolkien based the Western language, Quenya, on Finnish phonology and the Eastern language, Sindarin, on Welsh (Downing 1982; Noel, 1980; Phelpstead 2010). He eventually developed over fifteen Elvish languages, all based in part on these two proto languages.

While the Elvish languages were the most developed languages that Tolkien created, he also created languages for the other groups that inhabited Middle Earth. The most prevalent of these languages was Westron (also called *common speech*), dialects of which were spoken by most men and by hobbits. This also seems to be the common language in which all races communicated and is supposed to be the language from which English eventually descended. Other languages spoken by men included Rhovanion, Dalish, and Easterling, among others. All languages spoken by men were based mainly on Old English or other Germanic languages like Gothic and Old Norse or have been Anglicized (Foster 1978; Smith 1997). In fact, some of the men's names are neither derivatives nor representative of any of Tolkien's created languages, but are merely Anglicized. Nevertheless, these names were still created or selected by Tolkien.

Tolkien represented other groups as having developed or partially developed languages including Khuzdul, spoken by dwarves. Tolkien stated that he based Khuzdul on Semitic languages, since he considered the dwarves to be much like the Jews of the middle twentieth century, who had lived among a people that had greatly persecuted them (Drout 2007).Although Khuzdul was based on Semitic languages, many of the names of dwarves in Tolkien's works were "translated" into Westron for ease of pronunciation and because the dwarves liked to keep their language secret. These translated names were actually based on names for dwarves in two poems *Voluspa* and *Gylfaginning* (Allan 1978), which were based neither on Westron or Khuzdul, but are derived from Old Norse and then Anglicized. Many of Tolkien's original spellings are complex and included diacritical markings such as accents and diaereses. In this article they are not simplified for ease of transcription.

Other languages created were Entish (the language of the Ents or tree people), Black speech (the language spoken by Sauron, the primary antagonist of the *Lord of the Rings* trilogy) and Orkish, the language spoken by Orcs, a race created by Sauron. The later three languages were not based on one specific natural language or its phonology, but instead are "pidgins" or conglomerations of many languages created by Tolkien. In these last efforts Tolkien attempted to construct languages that were not based on Indo-European languages. For example, Black speech was a tonal language. However, it appears that none of the names of Orcs, Ents or other minor characters were necessarily based on these languages, but instead are meant to appear as translated equivalents.

Although the languages Tolkien created have elements of natural languages, it is more accurate to state that he based his languages only loosely on them. For example, the Sindarin name "Arwen," which is translated by Tolkien as meaning "noble maiden," can be traced to the morphemes *ar* (noble) and *wen* (maiden). These two forms may be based loosely on the Welsh words *arglwydd* (lord/noble) and *forwyn* (maiden) (Phelpstead 2010). Other names, such as *Legolas* are more difficult to trace, since Tolkien translated the name to "green leaf," even though this etymology does

not seem to relate to Welsh except in the sense that some sounds are similar (Phelpstead 2010). Still others seem to have been created by Tolkien without an obvious link to a clear etymology.

Many words from Westron were quite obviously borrowed from Old English (Smith 1997). For example, the name of the land of Sauron, *Mordor,* is the Old English word for *death* or *murder.* Similarly, the name Eomer, is composed of the Old English *eo* (horse) and *mar* (mare), meaning horse mare. Other words were more loosely derived from Old English such as *hobbit*, which Tolkien stated was taken from the old English *hol-bytla* (hole builder) and eventually was shortened to become hobbit (Noel 1980).However, neither *hobbit* nor *hol-bytla* were actually Hobbitish, but were actually *kuduk* and *kud-dukan* (see Appendix F in *The Lord of the Rings*). In fact, in this discussion of languages, Tolkien revealed that none of the Hobbit names in his works were actually Hobbitish, but were Anglicized or Westron equivalents (e.g., *Bralda-him* is Brandywine River and *Zaragamba* is Hobbitish for Oldbucks).

While Tolkien's characters had names derived from several "languages," four main groups are represented by a number of names sufficient for analysis: elves, men, hobbits, and dwarves. We added an "other" category to include Orcs, Ents, and Horses, the other groups for which a sizable number of names exist. Therefore, in this study we compared the phonological diversity of the names for the five groups to determine whether Tolkien was able to achieve a different phonoprint for each language. Even though these groups of names may actually be based on three languages according to Tolkien purists—the language of the Elves, Dwarves, and Westron (which are based on Celtic/Finnish, Old Norse, and Anglicized variants, respectively)—the fact is that Tolkien's names are primarily fictional (Foster 1978). He chose the names in order for them to appear distinct and to represent different groups. Our division into five categories and our use of the Anglicized versions of these names seems consistent with Tolkien's intent to create different names to represent different peoples, each with their own language.

The consonant and vowel inventory, phonotactics, and syllable structure of these languages are described in varying levels of detail depending on the language. Several analyses of these languages (as well as the others reported in this paper) are explained by Allen (1978), Salo (2004), and Tolkien (1971), among others. Tolkien provided an elaborate description of Quenya and Sindarin phonotactics and syllable structure. By contrast, he only provided the basic phonetic inventory of Khuzdal and explained that its phonotactics are similar to Semitic languages. While some of the languages of man are described in more detail than others, it appears that they all follow the basic phonotactics and syllable structures of Germanic languages (see Allan, 1978 for a description). Because we have limited understanding of what Tolkien wanted the phonotactic structure of some of the languages to be, we have opted to do an analysis of this structure based on the names we analyze below.

Although the Elf group used in this analysis involves two language groups (Sindarin and Quenya), the phonology and syllable structure of the two are similar enough (at least when we looked at the phonology and syllable structure of the names) that we combined them, particularly because Sindarin is a branch of Quenya (Foster 1978). In addition, Tolkien stated that some of the men's names were derived from Elvish such as Aragorn and Denethor (see Appendix F).

This study used 197 names from Tolkien that are not found in other sources, selected from an Internet list (http://lotr.wikia.com/wiki/List_of_characters) and verified using the online Encyclopedia of Arda (http://www.glyphweb.com/arda/). As few surnames or titles appear in the text, only first names were included. Names given to two or more characters were used only once in the study. Of the 197 names, 14 were discarded because they were names of beings for whom there were only a few representatives (dogs, ravens, goblins, etc.). Therefore, 183 names were used in the final analysis. Table 1 provides information about the names used from the five groups.

Table 1
Characteristics of the Names of the Five Groups

Name	N	Longest name (most number of phonemes)	Shortest name (least number phonemes)
Dwarf	23	Faldor (6)	Oin (2)
Elf	47	Celebrimbor (11)	Osse (3)
Hobbit	18	Bandobras (10)	Bob (3)
Man	74	Bladothin (9)	Tom (3)
Other (Ent, Orc, Horses)	21	Bregalad (8)	Arod (4)

Methodology

Three separate analyses were performed to determine whether Tolkien used different phonoprints for each of the five major language groups. These three analyses included phonotactic probabilities, frequency counts, and syllable structure.

Phonotactic Probabilities

Our first methodology examined whether the five groups of names differ in how "English-like" they are. To perform this analysis, we used the phonotactic probability calculator developed by Vitevitch, Luce, Charles-Luce, and Kemmerer (1997), accessed on the website <http://www.people.ku.edu/~mvitevit/PhonoProbHome.html>. This calculator determines the probability that each English phoneme occurs in a word position by comparing each word entered into the calculator to the frequency database of standard American English created by Kucera and Francis (1967). Each word is entered phonemically using the computer-readable transcription method called Klattese developed by Dennis Klatt (http://129.237.66.221/VLbrmic.pdf).

In these analyses, both phonemes and sequences of phonemes can be examined. For example, the vowel sounds /i/ and /I/ (which occur .03% and 6.23% of the time initially in words in the corpus) are more common in words initially in English than are /ai/ or /oi/ (which occur .002% and .04% of the time respectively). Similarly, the consonants /l/ (3.4%), /t/ (4.4%), /k/ (9.2%), and /n/ (2.3%) are more common in words initially than /j/ (.07%) or /w/ (2%) (Kessler & Treiman 1997). The calculator also reveals how frequently two sounds occur next to each other in different positions of the word. For example, it calculates the percentage of time /bi/ (.02%) occurs in words initially versus /br/ (.07%). These two-phoneme combinations are called *biphones* (bi-phonemes) by Vitevitch and colleagues. The calculator also determines the overall phoneme and biphone probabilities for an entire word. For example, for the names *Bilbo* and *Frodo,* the average probability that in English all the phonemes and biphones would come in the order they do is higher for Frodo (.04334) than for Bilbo (.0389). Phonotactic probability does not measure how common a word or name is. Bilbo's name may be used more frequently than Frodo's name in Tolkien's works. Phonotactic probability deals only with prevalence of the sound sequence.

Two analyses were done using the phonotactic probability calculator. We first examined the overall probability of having all the phonemes and biphones in a name occur in the order that they do. We next examined the probability that the sounds in the word occurred in the ordinal position that they do, looking at each position separately. We performed this last analysis only for the first four positions, since most of Tolkien's names had at least four phonemes.

Frequency Counts

The second method of analyzing the names was by examining the frequency with which phonemes were used within each of the five name groups. This analysis was performed to determine whether some of the groups used different phonemes to a greater or lesser degree than did other groups. Two such frequency analyses were performed. The first was to determine the number of times each sound was used in the names of the five groups. For example, there are 74 human names which are composed of 458 different sounds. The phoneme /b/ appears in the names 25 times, meaning that this sound appears approximately 5% of the time (25/458). Similarly, dwarf names have 100 different sounds, and the phoneme /b/ appears 7 times, or 7% of the time (7/100).

In addition, we also analyzed how often certain sounds appeared in the first four positions of the names. For example, in the names of men /b/ appears as the first sound in 22 of the 74 names, or approximately 29% of the time. In Dwarf names /b/ appears as the first sound 6 times in the 23 names or 26% of the time. As with the probability counts, this analysis was done only for the first four ordinal positions of the name.

Syllable structure

The final analyses determined whether there are significant phonotactic differences and syllable structure differences across the five types of names. Two analyses were performed. The first was to determine what percentage of syllables used in the names follow the most common patterns of syllable structure found across the world's languages: V, CV, CVC, CCV, CCVC, CVCC, CCVCC, VC, and VCC. For example, of all the syllables in dwarf names, 46% are CV and 35% are CVC. Of all the syllables in men's names, 33% are CV and 29% are CVC.

For most names, the syllables were easy to divide. However, in words based on English, determining syllable structure is often difficult because of ambisyllabicity (cases in which a sound can be in the coda position of one syllable and the onset in an adjacent syllable as in [hæp.pi]). To determine syllable patterns we followed the maximal onset principle, which states that languages tend to maximize onsets, not codas. Because of this, in names such as "Balin" or "Borin," we divided the names ba.lin and bo.rin.

The second analysis was to examine what types of onset and codas were used and were acceptable and how they differed across the five name groups. For this analysis, in addition to using the maximal onset principle, we also assumed that onsets were composed of consonant clusters that are acceptable for the natural languages upon which Tolkien's languages were based. For example, we assumed that "Bilbo" was composed of two syllables "bil" and "bo" since "lb" is not an acceptable onset in English or in any of the languages upon which Tolkien based his work (Welsh, Finnish, Germanic, Semitic) or in any of the languages he created (Sindarin, Quenya, Westron, etc.). In this analysis we compared the most common onsets and codas used in the names and looked for specific patterns.

Results

The first analyses of this study examined whether the five groups of names differed in terms of their likeness to English. This was done by comparing the average phoneme and biphone probabilities for each word in the five name groups. See Table 2 for a comparison of average phoneme and biphone probabilities as well as average name length for the five groups of names. In this analysis, we also examined the average number of phonemes in each word. We hypothesized that the five groups

would differ in how "English-like" they were since some of the language groups were based on Old English and other Germanic languages (Hobbit, Man) and others were based on non-Germanic languages (Elvish, Dwarf, Other).

A one-way MANOVA was used to compare the five name groups within the three-dimensional space defined by: overall name phoneme probability, overall name biphone probability, and length of name. The five name groups differed significantly on all four multivariate tests (Wilks' Lambda = .80429, p=0.0001, multivariate R^2=.196; Pillai's Trace = .20203, p=0.0002; Hotelling-Lawley Trace = .23552, p<.0001; and Roy's Greatest Root = .16511, p<.0001). These significant multivariate results were followed up with three one-way ANOVAs. The one-way ANOVAs revealed no significant differences among groups for overall name phoneme probability (F(4, 178) =1.46, n.s.) or for biphone probability (F(4,178) = 1.38, n.s.), but the results did reveal significant differences among the groups for name length (F(4,178) = 7.88, p< .0001, R^2=.151), indicating that most of the multivariate significance was due to name length.

Post hoc analyses were used to determine which group comparisons were responsible for the effect on name length. Because the name groups differed markedly in size, the Scheffe's (1953) method of post-hoc comparisons was used. As can be seen in Table 2, the dwarf names are noticeably shorter in length than the other four groups. This was evident in the post hoc comparisons, where the four comparisons involving dwarf names were statistically significant, with the four t values ranging from 3.935 (p=0.0049) to 5.281 (p<.0001). The other six post hoc comparisons were not significant.

Table 2
Average Phoneme Probabilities, Average Biphone Probabilities, and Length of Words for Each of the Five Groups of Names

	Phoneme AVG	Biphone AVG	Length
Dwarf	.046	.0038	4.348
	Range from Balin (.0618) to Azaghal (.0174)	Range from Nori (.0094) to Oin (.0001)	
Elf	.044	.0032	6.191
	Range from Sauron (.0792) to Osse (.0122)	Range from Indis (.0126) to Ingwe (.0004)	
Hobbit	.042 Range from Peregrin (.0654) to Angbor (.0203)	.0030	6.220
		Range from Barliman (.0060) to Bungo (.0010)	
Man	.044	.0036	6.203
	Range from Bill (.0737) to Olwe (.0184)	Range from Saruman (.0101) to Olwe (.0003)	
Other	.038	.0025	6.095
	Range from Felarof (.0566) to Azag (.0098)	Range from Wandlimb (.0052) to Azag (.0002)	

Note: Ranges from most to least English-like are given in parentheses.

We ran a similar MANOVA analysis on phonemic probabilities for each of the first four ordinal positions in each name. The five name groups differed significantly on all four multivariate tests (Wilks' Lambda = .80294, p=0.0020, multivariate R^2=.197; Pillai's Trace = .20994, p=0.0021; Hotelling-Lawley Trace = .22966, p=0.0022; and Roy's Greatest Root = .10830, p=0.0006). In the follow-up univariate tests, only phoneme position three had significant effects ($F_{(4,170)}$ = 3.20, p=0.0146, R^2=.070). Figure 1 displays a comparison of average phoneme probabilities across the five Tolkien name groups at each of the first four ordinal positions. The Scheffe post-hoc tests revealed only one significant comparison between the elf group of names and the mixed group of names (t=3.130, p=0.0482). In Figure 1 it can be seen that the largest difference among the profiles is in position three, with the elf group of names having the highest phonemic probability (0.0574) and the mixed group of names having the lowest (.0339).

Figure 1. A comparison of the phoneme probabilities across the five Tolkien name groups at each ordinal position.
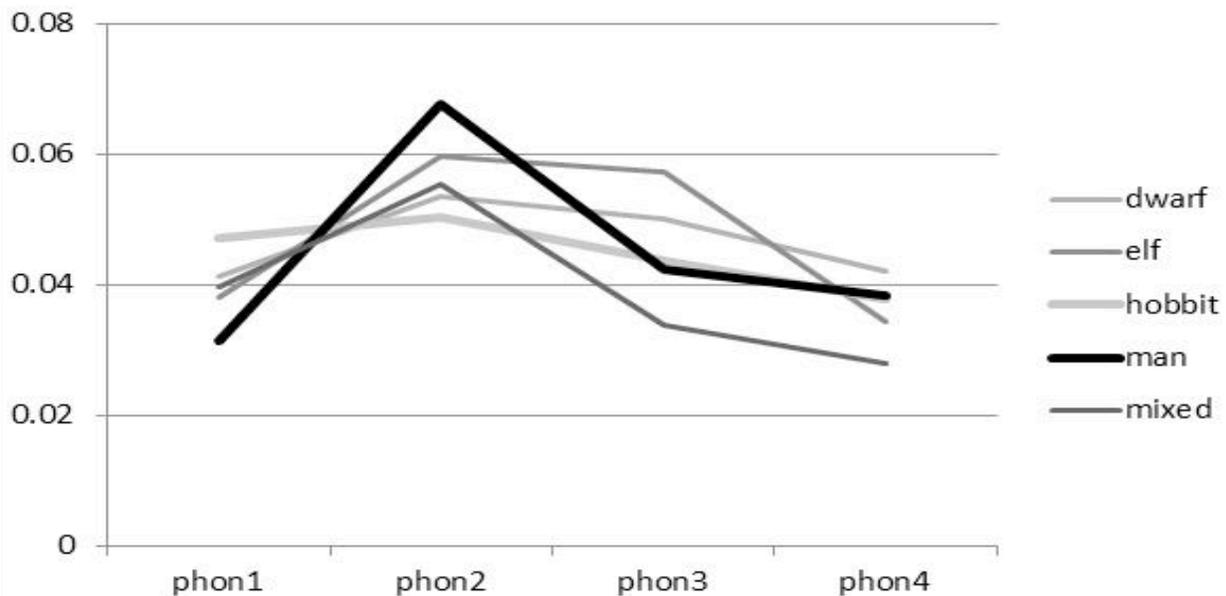
Frequency Analyses

Our next group of analyses examined the overall frequency with which certain sounds were used in the five groups of Tolkien names. The first analysis revealed whether the groups differed when considering which phonemes were used in the names and how often they were used. Table 3 provides the total number of phonemes in each group of languages, along with the total number of different phonemes, and the ten most common phonemes used in each of the five groups. When the names are considered as a whole, Tolkien used only 26 of the 44 possible phonemes in English, and on average each group of names includes about 21.8 of these phonemes. The most common sounds used across the five name groups are /ɾ/, /n/, /e/, /a/, /l/, /I/, /o/, /x/, /b/, and /d/.

Table 3
Total Number of Phonemes, Total Number of Different Phonemes and Ten Most Common Phonemes Used for Each of the Five Name Groups

	Total number of phonemes	Total number of different phonemes	Ten most common phonemes (in order)
Dwarf	100	18	ɪ, n, o, r, l, b, ɑ ,d, m ,f
Elf	291	21	n, ɛ, r, l, ɪ, ɑ æ, o, d, g
Hobbit	112	24	b, ɑ, r, o, d, l, x, e , m , n
Man	458	24	r, ɛ, n, ɑ, x, o, l, b, θ, ɪ
Mixed	128	22	ɑ, r, g, l, k, n, ∫, æ, ɪ, b

Our first analysis compared the number of times a sound appears in a particular group of names to the average number of times it is used in the whole set of names. As an example, Figure 2 demonstrates the differences between the average number of times a particular phoneme appeared across all the names and the number of times they appear in the Dwarf names. In other words, we compared the number of times /r/ was used in dwarf names compared to the average number of times /r/ was used across all the groups, the number of times /r/ was used in man names compared to the average number of times it is used across the groups, and so on through all the phonemes and all the groups.

Figure 2. The frequency with which phonemes are used on average across the five name groups and the frequency with which phonemes are used for the Dwarf names
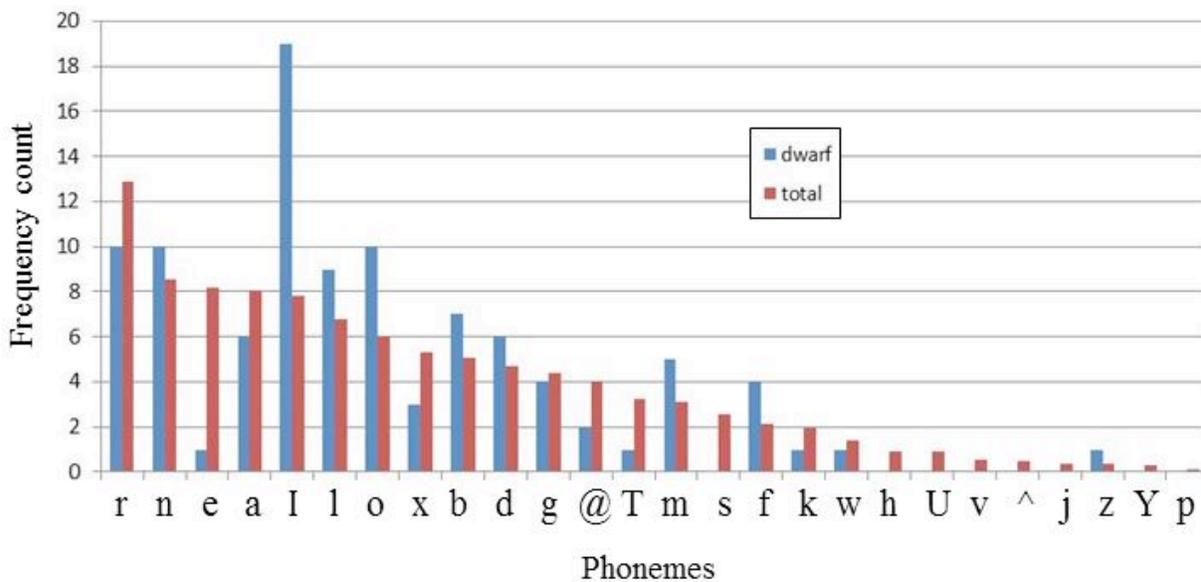
Because so many analyses were performed, we used a Bonferroni correction and set the *p* value to .00001. Of the 130 z tests of significance comparing the proportion of each phoneme for each name group with the corresponding proportion over all of the Tolkien names, only eight are significant

when considered individually, but only one phoneme in one language group remains significant when adjusted for multiple testing (see Table 4). The phoneme /I/ was used more in dwarf names than in the other 4 groups.

Table 4
The Eight Strongest Results of 130 z Tests Comparing the Proportions of Each Phoneme in the Five Name Groups to the Proportions of That Phoneme in the Total Group

Name Group	Phoneme	z-test value	Direct alpha	Bonferonni adjusted significance level
Dwarf	I	3.792	.00007	Significant
Hobbit	b	2.702	.00344	Not significant
Hobbit	ʌ	1.660	.04851	Not significant
Hobbit	z	1.916	.02770	Not significant
Hobbit	ɑɪ	2.234	.01276	Not significant
Hobbit	p	1.871	.03069	Not significant
Other	s	2.055	.01995	Not significant
Other	k	3.114	.00092	Not significant

For the next frequency analysis, we examined whether each of the five groups tended to use the same phonemes for each of the first four ordinal positions in the personal names. Figure 3 provides the percentage of names from each of the five name groups that is accounted for by the two most common phonemes in each of the four ordinal positions (the top two phonemes are given above the bars). In all five groups, the percentage of names accounted for by the two most common phonemes was on average 41%. Moreover, for most of the groups, the most common phonemes were the same: /b/, /r/, /n/, /l/, / æ/, /ɛ/, and /ə/. A simple perusal of the consonants used as the most common first few sounds of the names in the five groups reveals other commonalities. For example, the sound /b/ is the most common first phoneme for three of the five groups (dwarf, hobbit, man).Moreover all of the groups of names have either /r/ or /l/ as one of the most common phonemes in the second or third position.

When we compared the four groups, the chi square value was not significant ($\chi^2(12)=$ 5.2074, $p=.9507$). In other words, there is no evidence that the five name groups differ in phonemes used in each of the four ordinal positions.

Figure 3. Bar graph showing the percentage of names from each of the five sources accounted for by the two most common phonemes in each of the four ordinal positions (top two phonemes are given above the bars
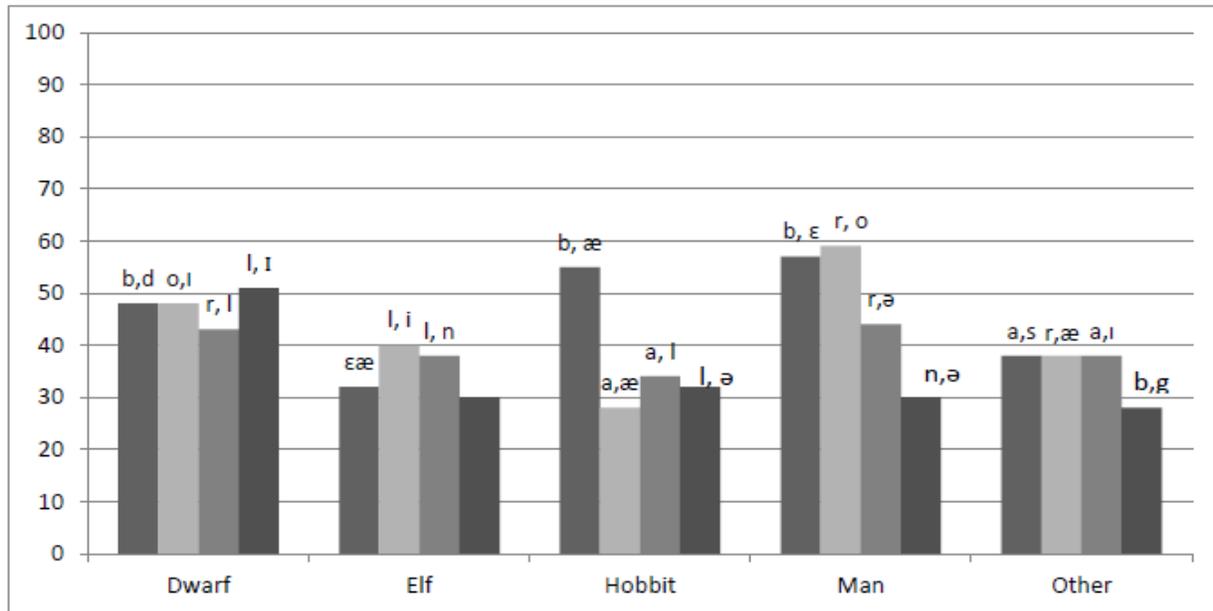
Syllable Structure and Phonotactics

Our final set of analyses compared the syllable structure and phonotactics of the five name groups. For the first of these analyses, we compared the syllable structure for each of the five name groups. Table 5 provides the percentage of each of the nine types of syllables used for each of the groups of names from the most simple syllable (V—or a syllable composed of just one vowel sound) to the most complex CCVCC (with two consonants in the onset and two in the coda). Chi-square analyses were performed to determine if any of the groups differed from each other in the frequency with which any of the syllable types were used. (Because of the number of analyses, a Bonferonni correction was employed and the p value was set at .005). The results of these analyses revealed that there was no difference among the groups (all χ^2's(4) <11.26, all p's > .02) except for VC syllables (χ^2(4) =24.28, p=.0001). Elf names were more likely to use VC syllables than were the other four groups (as in *Elrond*). (Note: No analysis was performed on the CCVCC syllable structure since most of the groups did not have any syllables that followed this pattern).

Table 5
Percentage of Each Syllable Type for the Five Name Groups

	N syll	V	CV	CVC	CCV	CCVC	VC	CVCC	VCC	CCVCC
Dwarf	37	5%	46%	35%	3%	8%	2%	0	0	0
Hobbit	49	14%	35%	31%	4%	6%	4%	4%	2%	0
Elf	165	8%	37%	33%	1%	2%	18%	1%	0	0
Man	204	16%	33%	29%	2%	2%	9%	7%	.04%	.04%
Other	53	8%	26%	36%	7%	9%	3%	7%	1%	0

Our second analysis examined what consonant clusters were used for the onsets and codas for each of the five name groups (see Table 6). Three patterns from the data are clear: First, Tolkien seems to favor onset consonant clusters that combine voiced plosives (/g, d, b/) with liquids (/l, r/) or glides (/w/). Only the "other" category has combinations other than these (/sn/, /kw/, /sk/). Second, most sounds in coda position are liquids (/l, r/) or nasals (/n, m/). In fact, single consonant codas in the Dwarf category are only nasals and liquids. For the hobbit, elf, and man names, most of the single consonant codas are the same (87%, 76% and 85%, respectively). Again, only the other category has final consonants that differ from the other groups. Finally, coda consonant clusters are rare for any of the five name groups, but they are also usually composed of nasals and liquids, along with an alveolar plosive (i.e., /nt/, /rd/, /nd/). The two groups that differ from this pattern are man names (/lf/, /st/) and the names in the other category (/ks/, /rk/). In both cases, however, these exceptions are rare. Considering all the onset and coda consonant clusters available in English, Tolkien is only using a small subset of these possibilities.

Table 6
Descriptions of Onsets and Codas in the Syllables of the Names of the Five Groups
(number of each type is given in parentheses)

	Onset consonant clusters	Single Consonant Codas	Coda consonant clusters
Dwarf	dw (1)	l (2)	None
	gl(1)	m (3)	
	dr(1)	n (5)	
		r(4)	
Hobbit	dr(1)	b (1)	nd(1)
	fr(1)	l (6)	nt (1)
	bl(1)	n (7)	rd (1)
	br(1)	r (2)	

		gr(1)	z(1)	
Elf	br(2)	l (17)	nd (1)	
	dr(1)	m (3)	rn (1)	
	gr(1)	n (34)		
	gl(1)	ŋ (2)		
		r (12)		
		s (5)		
		θ (3)		
Man	bl (1)	d (3)	ld (1)	
	br (5)	g (2)	lf (1)	
	dr (1)	k (1)	lm (1)	
	gl (1)	l (12)	nd (2)	
	gr(1)	m (2)	nt (1)	
		n (21)	rd (1)	
		r (30)	rm (1)	
		s (4)	rn (5)	
			rθ (1)	
			st (1)	
Other	br(1)	b (1)	ks (1)	
	gl(1)	d (3)	nd (1)	
	gr(1),	f (2)	nt (1)	
	kw(1)	g (3)	rk (1)	
	sk(1)	k(5)	rd (1)	
	sn (2)	m(2)		
	tr (1)			

n (3)
r (1)
ʃ (1)
θ (1)

Discussion

The purpose of this study was to determine whether Tolkien was able to differentiate the phonotactics (i.e., have a different phonoprint) for each of his five major naming groups. The results of this study seem to suggest that Tolkien was not able to differentiate the phonotactics of each group, but instead has a phonoprint that he used to create names, regardless of which language group the names originated from. These results and their significance are discussed in order of the research questions in more detail below.

Phonotactic Probability

In our first analysis, we examined whether the five groups of names differed in their phonotactic probabilities and determined that there was no difference across the groups in how "English-like" they were, at least in terms of their overall phoneme and biphone probability. In fact, the only difference we found across the groups was that Dwarf names were found to be significantly shorter (in terms of average number of phonemes) than the names of the other groups. The reason for this difference may have been that Tolkien tried to cast Dwarfs as outsiders similar to Jews during his time period (Drout 2007) and because he based these names on dwarf names found in the poems *Voluepa* and *Gylfaginning* (Allan 1978). It may also be that a difference in name length was a conscious effort to differentiate Dwarves and match their names to their short stature.

When we examined the phonotactic probability for each of the first four ordinal positions, we found that the only difference between the five groups was that the Elf names and the other names differed in terms of the probability of the third phoneme. Elf names had phonemes with a highest phoneme probability and the names in the other group with the least phoneme probability. A perusal of the third phonemes of each of these groups demonstrates that the most common third phonemes in the other name group were /æ/ (.0283) and /g/ (.018). By contrast, the most common third phonemes for the Elf name group were /l/ (.073) and /r/ (.078). In other words, the names in the other name group are composed of less common sounds and sound combinations (i.e., Smaug, Bolg, Roac). As demonstrated by the syllable structure analysis as well, the other names seemed to almost break Tolkien's phonoprint. It's possible that because these names represent characters who are unique, Tolkien also tried (and almost succeeded) in making their names distinct as well.

Although not statistically significant, we also found that man names differed from the other four groups in how English-like they were. In particular, they were less English-like in the first position and most English-like in the second. This may be due to the fact that Tolkien often started man names using phonemes that were more common in OE than they are now. For example, many of the names started with /θ/ (Theoden, Theodred) or with vowels common to OE names such as /æ/ and /ɛ/ (Asgan, Eomer). This may have been a conscious attempt by Tolkien to make man

names appear less like names in modern English. It is also possible that the man names were more English like in the second position than the other four groups because for man names that started with a vowel, the second position was usually a consonant.

What is also striking about this analysis is that other than these two differences, the names across the five groups were very similar in how English-like they were, demonstrating that Tolkien's phonoprint was consistent across the five name groups.

Frequency Analysis

Our next group of analyses investigated how frequently certain phonemes were used across the five name groups. For the first of these, we examined how frequently each phoneme was used by each of the groups in comparison to the average. Our results revealed that for the most part the five groups used the same phonemes an equal amount of time. Only the Dwarf names differed from the average and only for one phoneme (/ɪ/).

Even with careful planning to create new names, this analysis revealed that Tolkien tended to use the same phonemes over and over: /b/, /r/, /n/, /l/, /a/, / æ /, / ɛ /. By contrast, he rarely used some phonemes (/t/, /p/, /f/), and did not use some at all (/tS/), despite their common use in English. This finding again highlights that Tolkien had a particular disposition to choose some phonemes over others, and to use them in a different frequency than they are used in English. He used his phonoprint consistently across the five name groups.

When we examined the frequency with which different phonemes were used in the first four ordinal positions of the names, we found that most of the names used only a few of the same phonemes for the first four sounds. In fact, there were no significant differences in frequency of sounds used. The same phonemes were used in all the five name groups: for example, /r/ (man: *Aragorn*, hobbit: *Frodo*, other: *Grishnak,* and dwarf: *Ori*), and /o/ (man: *Boromir*, elf: *Lorien*, dwarf: *Nori*, and other: *Roac*) appeared as the second position for many of the groups' names. In most cases, only two phonemes accounted for strikingly half of the sounds in each of the ordinal positions. What this demonstrates is that Tolkien uses the same sounds over and over again in all of the first four positions of the five name groups.

Syllable Structure and Phonotactics

Our final analyses examined the syllable and phonotactic structure of the five name groups. In this analysis we found that there was little difference in the types of syllable structure used in each of the languages. In fact, there was only one difference: Elvish names had more VC syllables than any other group. This is likely because so many of the elves are named with "El" (*Elrond, Elwe, Elwing, Eldarion*) or with "Ar" (*Arwen, Arminas, Aranwe, Arien*). In addition, when we examined the onset and codas of the syllables, it was apparent that Tolkien used the same onset and coda combinations over and over again. He favored voiced obstruent-liquid combinations for onset (/br/, /gr/, /gl/, /dr/) and nasals and liquids for codas (/n/, /m/, /l/, /r/, /nt/, /rn/, etc). Only the "other" group showed occasional variation from this pattern.

Determining that all the name groups follow approximately the same phonotactics is significant, since Tolkien claims to have had distinct phonotactics for his different languages and to have based them on very different modern languages. Instead, most of the names follow syllable structure that follows a subset of possibilities in English or Germanic languages.

Tolkien's Phonoprint

The results of this analysis suggest that Tolkien's phonoprint is comprised mostly of these sounds (/b/, /r/, /n/, /l/, /a/, / æ /, / ɛ /) and the onsets /br/, /gr/, /gl/, and /dr/ and the codas/n/, /m/, /l/, /r/, /nt/, and /rn/. While not all of the names he included in his work followed this pattern, a surprisingly large number do, despite his conscious effort to invent languages that differed from each other.

Conclusions

Based on the findings of this study, it appears that even talented authors cannot escape their own phonoprint (conscious or subconscious phonological tendencies) even when they purposely try to create or represent different languages. Such findings suggest that phonoprints may be another method of author identification. Our results are tentative since they are based on only one author, however linguistically talented he may have been. More studies need to be done using more authors to verify this preliminary result. In addition, more extensive research needs to examine Tolkien's original names and words that he claimed were authentic in his created languages (such as Azanulbizar in Khudzul, Ar-Adunakor in Westron and Brandagamba in Hobbitish) rather than the Anglicized versions that appear in his texts and were examined in this study.

References

Aceto, Michael. "Ethnic Personal Names and Multiple Identities in Anglophone Caribbean Speech Communities in Latin America." *Language and Society* 31 (2002): 577-608.

Allan, James. *An Introduction to Elvish: And to the other tongues and proper names and writing systems of the third age of the western lands of middle-earth as set forth in the published writings of Professor John Ronald Reuel Tolkien.* Somerset: Bran's Head Books, 1978.

Lo Bianco, Joseph. "Invented Languages and New Worlds." *English Today* 78 (2004): 8-18.

Black, S. & Wilcox, B. "Sense and serendipity: Some ways fiction writers choose names." *Names: A Journal of Onomastics* 59 (2011): 152-163.

Croft, D. J. "Book of Mormon 'Wordprints' Reexamined." *Sunstone* 6 (1981): 15-22.

Downing, Angela. 1982. "From Quenya to the Common Speech: Linguistic Diversification in J. R. R. Tolkien's *The Lord of the Rings." Revista Canaria de Estudios* 4 (1981): 23-42.

Drout, Michael D. C. (ed). *J. R. R. Tolkien Encyclopedia: Scholarship and Critical Assessment.* New York: Routledge Taylor and Francis Group, 2006.

Foster, Robert. *The Complete Guide to Middle-earth: From The Hobbit to The Silmarillion.* New York: Ballantine Books, 1978.

Fudge, Eric and Linday Shockey. "The Reading Database of Syllable Structure." *Linguistic Databases,* ed. by John Newborne, pp. 93-102. Stanford: CSLJ Publication, 1998.

Grieve, Jack. "Quantitative Authorship Attribution: An Evaluation of Techniques." *Literary and Linguistic Computing* 22 (2007): 251-270.

Hilton, John L. "On Verifying Wordprint Studies: Book of Mormon Authorship." *BYU Studies* 30 (1990): 89-108.

Holmes, David I. "Authorship Attribution." *Computers and the Humanities* 28 (1994): 87-106.

Hough, Carole. " Towards an Explanation of Phonetic Differentiation in Masculine and Feminine Personal Names." *Journal of Linguistics,* 36 (2000): 1-11.

Iqbal, Farkhund, L. A. Khan, Benjamin C. M. Fung, and Mourad Debbabi. E-mail verification of forensic investigation. In Proceedings of the 25th ACM SIGAPP Symposium on Applied Computing, pp. 1591-1598. Sierre, Switzerland: ACM Press, 2010.

Kucera, Henry and W. Nelson Francis. *Computational Analysis of Present-day American English.* Providence: Brown University Press, 1967.

Lamaj, Artur and Valter Memishaj. The Use of Personal Names in Albanian. *Cahiers Balkaniques* 32 (2001): 31-37.

Leglise, Isabelle and Bettina Migge. "Language-naming Practices, Ideologies, and Linguistic Practices: Toward a Comprehensive Description of Language Varieties." *Language in Society* 35 (2006): 313-339.

Lombard, Carol G. "The Sociocultural Significance of Niltsitapi Personal Names: An Ethnographic Analysis." *Names* 59 (2011): 42-51.

Luyckx, Kim and Walter Daelemans. "The Effect of Author Set Size and Data Size in Authorship Attribution." *Literary and Linguistics Computing,* 26 (2011): 35-55.

Mhlambi, Innocentia J. "Acts of Naming: The Detective Plot in Masondo's Fiction." *South African Journal of African Languages* 3 (2007): 128-141.

Miranda-Garcia, Antonio and Javier Calle-Martin. "Function Words in Authorship Attribution Studies." *Literacy and Linguistic Computing* 22 (2007): 49-66.

Morton, Andrew Q. *Literary detection.* New York: Scribner's Sons, 1979.

Nemer, Julie F. "Phonological Stereotypes and Names in Tenne." *Language in Society* 16 (1987): 341-352.

Noel, Ruth S. *The Languages of Tolkien's Middle-Earth.* Boston: Houghton Mifflin Company, 1980.

Nummila, Kirsi-Maria. "Occupational Designations Derived from Nouns + -(uri)- Among our Personal Names." *Virittaja* 111 (2007): 543-566.

Phelpstead, Carl. *Tolkien and Wales: Language, Literature, and Identity.* Cardiff: University of Wales Press, 2010.

Rubio Orecilla, Francisco J. "Celtic *Kuono-* in Hispano-Celtic Personal Names." *Beitrage zur Namenforschung* 41 (2006): 399-410.

Rymes, Betsy. "Naming a Social Practice: The Case of Little Creeper from Diamond Street." *Language in Society* 25 (1996): 237-260.

Salo, David. *A Gateway to Sindarin: A Grammar of an Elvish Language from J. R. R. Tolkien's Lord of the Rings.* Salt Lake City: The University of Utah Press, 2004.

Scheffe, Henry A. "A Method for Judging all Possible Contrasts in the Analysis of Variance." *Biometrika,* 40 (1953): 87-104.

Smith, Arden R. *Germanic Linguistic Influence on the Invented Languages of J. R. R. Tolkien.* Unpublished doctoral dissertation. UC Berkeley, 1997.

Stefanov, Pavel. "Bulgarian Personal Names of Romanian Origin in a Manuscript from 1720." *Spostavitelno Ezikoznaie/Contrastive Linguistics,* 16 (1991)*:* 27-30.

Superanskaya, Aleksandra V. "Russian Personal Names." *Folia Onomastica Croatica* 8 (1999): 191-200.

Suzman, Susan M. "Names as Pointers: Zulu Personal Naming Practices." *Language in Society* 23 (1994): 253-272.

Tolkien, J. R. R. "Quenya Phonology." *Parma Eldalamberon 19,* 18.01.11.

Verstraten, Freya. 2006. "Naming Practices Among the Irish Secular Nobility in the High Middle Ages." *Journal of Medieval History* 32 (1971): 43-53.

Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce, and David Kemmerer. "Phonotactics and Syllable Stress: Implications for the Processing of Spoken Nonsense Words." *Language & Speech* 40 (1997): 47-62.

Vitevitch, Michael and Paul A. Luce. "A Web-based Interface to Calculate Phonotactic Probability for Words and Nonwords in English." *Behavior Research Methods, Instruments & Computers* 36 (2004): 48-487.

Wenzel, Walter. "Interrelations Between Lower Lusatian Placenames and Personal Names." *Onoma* 36 (2001): 165-179.

Wilcox, B., Brown, B. L., Baker-Smemoe W., Eggett, D., Hoskisson, P. Y., and Black, S. "Identifying Authors by Phonoprints in Their Characters' Names: An Exploratory Study." *Names: A Journal of Onomastics* 61 (2013): 104-125.

Zheng, Rong, Yi Quin, Zan Huang, and Hsinchun Chen. "Authorship Analysis in Cybercrime Investigation." *Intelligence and Security Informatics Proceedings* 2665 (2003): 59-73.