Tolkien's Phonoprint in Character Names Throughout His Invented Languages

BRAD WILCOX

Department of Ancient Scripture, Brigham Young University, Provo, Utah, USA

BRUCE L. BROWN

Department of Psychology, Brigham Young University, Provo, Utah, USA

WENDY BAKER-SMEMOE

Department of Linguistics, Brigham Young University, Provo, Utah, USA

TIMOTHY G. MORRISON

Deparment of Teacher Education, Brigham Young University, Provo, Utah, USA

J. R. R. Tolkien claimed to create names based on sound symbolism and invented languages. Previous studies revealed that Tolkien had his own phonoprint (tendency to use the same phonemes repeatedly for character names) regardless of his claims that they reflected different language origins. However, this research examined phonemes and syllables rather than names as whole units. This study compared 183 names from Tolkien's works and 100 names from the nineteenth Century US Census. The name recognition software linked names with known languages or provided a "generic" category when no match was found. Results showed no significant differences between the generic listings when language groups were compared. It seems Tolkien successfully created unique names, but could not escape his own phonoprint among the generic names. When looking at names from non-generic responses, it appears Tolkien was able to differentiate into three language groups, but not with the precision he claimed.

Much research has been done describing the importance of names in literature (Algeo 1982; Black and Wilcox 2011; van Dalen-Oskam 2013). Character names have been found to have important functions in fiction: to provide a sense of "normalness" or "realness" (Debus 2002); to foreshadow the character's importance in the plot; and to portray a character's personality and express that a character is good or evil. For example, Croft (2009) explained that in Rowling's works, good characters are often afraid to use evil characters' names – choosing instead to use euphemisms such as "The Dark Lord" or "He-Who-Must-Not-Be-Named." Such functions are especially important in fantasy literature since writers often attempt to create a believable world (Algeo 1982; Croft 2009; Robinson 2013a).

One fantasy author who has received much research attention is J. R. R. Tolkien (Robinson 2013b; Smith 1997). Most of the studies investigating Tolkien's character names focus on their etymology, and attempt to determine whether or not they were derived from real or invented languages. Even Tolkien himself published several articles explaining the origin and reasoning behind the names that he chose for his characters (Tolkien 1971, 2006). He claimed many of his names were derived from sound symbolism and invented etymologies.

Past research

Researchers have noticed that many fantasy authors have used sound symbolism and phonological patterns when creating character names (Algeo 2001; Burelbach 1982; Robinson 2013a). Indeed, Robinson (2013a) reviewed several critics' descriptions of how sounds may call forth different emotions. One example is Jesperson's (1964) discussion of high front vowels indicating light (*gleam, glimmer, glitter)* and back vowels indicating darkness (*gloom*). Names created using such sound symbolism have been shown to successfully elicit words and images related to the sounds. For example, Croft (2009) postulated that both Rowling and Tolkien chose sounds like "sau- and sly- … calling forth imagery of snakes and cold-bloodedness" (p. 162).

Tolkien's sound symbolism

Tolkien was a professor of Anglo Saxon at Oxford who was fascinated with the structure, historical development, and relationships among languages. Given his background in philology and his love of languages, using sounds to invoke images was a vital part of his creative process as a writer. He called his use of certain sounds for his characters "phonetic fitness," which he seemed to relate not only to how a sound symbolized a character's personality, but also to Tolkien's own love for individual and combined sounds (Robinson 2013b; Smith 1997). Tolkien (2006) himself explained the following:

> If I were pressed to give any example of a feature of this style, not only as an observable feature but a source of pleasure to myself, I should mention the fondness for nasal consonants, especially the much favoured *n*, and the frequency with which word-patterns are made with the soft and less sonorous w and the voiced aspirants *f* and *dd* contrasted with the nasals: *nant*, *meddiant*, *afon*, *llawenydd*, *cenfigen*, *gwenyn*, *cranfanc*, to set down a few at random. (193-194; as cited in Robinson 2013b, 69)

Tolkien's languages

Tolkien often based his character names on the sounds found in real languages. Obviously, Tolkien preferred some sounds over others. His choices were deliberate, not haphazard. For example, he often used morphology of English, such as the morpheme *mor,* which is related to death, and was included in both *Morgath* and *Mordor.* In other cases, Tolkien based his character names on languages he had created. Other names were based on languages Tolkien insisted had unique phonological properties not related to English. For example, a stem from one of his invented languages, Quenya, is *ar,* which means "high" or "noble." From this stem come the names of Aragorn and Arwen, both royalty in his mythology (Noel 1980). Tolkien did not create languages to fit Middle Earth, rather Lo Bianco (2004) claimed that Tolkien created Middle Earth so his languages would have a place to roam free.

One cannot categorize the languages and the unique character names of Tolkien without a basic understanding of his mythology. According to Tolkien mythology, at the beginning of time were the Valar, from whom the Western and the Eastern Elves descended. Tolkien based the invented Western language, Quenya, on Finnish phonology and the invented Eastern language, Sindarin, on Welsh (Downing 1982; Noel 1980; Phelpstead 2010). He provided elaborate descriptions of the phonotactics and syllable structure of these languages and created over 15 Elvish languages, all based in part on these two proto languages (Salo 2004; Tolkien 1971). These Elvish languages were his most developed, however he also created languages for other groups. Westron (also called *common speech*) gave birth to the dialects spoken by hobbits and most men. Tolkien used the word *men* to describe humans, both male and female. Throughout this article we used Tolkien's original term. Rhovanion, Dalish, and Easterling are other languages spoken by men that were based primarily on Old English or other Germanic languages like Goth and Old Norse (Allan 1978; Smith 1997).

Other groups were said by Tolkien to have only partially developed languages. Khuzdul, spoken by dwarves, was an example of one such language. Tolkien explained that he based Khuzdul on Semitic languages since he considered the dwarves to be much like the persecuted Jews of the middle twentieth century (Drout 2006; Tolkien 1971). Many dwarf names were based on two poems, the *Voluepa* and *Gylfaginning*, and "translated" into Westron (Allan 1978). Other languages created by Tolkien were Black Speech, the language spoken by Sauron, the primary antagonist of the *Lord of the Rings* trilogy, Orkish, the language spoken by orcs, a race created by Sauron, and Entish, the language of the ents, or tree people. These three languages were not based on one specific known language, but instead are composites of many languages – pidgins so to speak – invented by the author who specifically did not want them to sound like Indo-European languages. For this reason Tolkien made Black Speech a tonal language. Using sound symbolism and his invented languages, Tolkien created names for characters in Middle Earth that were meant to be unique. They were also meant to be different according to each language. Was he successful?

Purpose

Previous research has suggested that Tolkien had his own phonoprint – a tendency to create names using the same phonemes and syllable structures regardless of the languages from which they derived (Smemoe et al. 2014; Wilcox et al. 2013). However,

these studies were conducted looking only at phonemes and syllable structures and not names as whole units.

The primary purpose of the current study was to compare names as whole units from five different language groups (dwarf, elf, hobbit, man, and other) as well as male names from the nineteenth century census using IBM name recognition software, the same software used by airport personnel to identify probable language backgrounds of passengers by their names. Although the software was created for modern languages and Tolkien's work is supposedly based upon ancient languages, the software provides a "generic" category when there is no match between a name and any known language. It is the frequency of that generic description that was used as a measure of the uniqueness of Tolkien's character names as whole units and how the names varied from one language to another. A secondary purpose of this study was to report the various languages the IBM software linked to the Tolkien names and determine how closely these matched the languages Tolkien intended.

Research questions

The following three research questions guided this investigation: (1) What is the frequency of the generic category in name recognition software within Tolkien's various invented languages?; (2) What is the frequency of modern languages linked with Tolkien's names by name recognition software?; (3) How closely do reported languages match Tolkien's intended language sources?

Methodology of current investigation

This study began with 197 names from Tolkien's writings that we 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/). Only first names were included since few surnames or titles appeared in the text. Names given to two or more characters were used only once in the study.

The 197 names came from the four main groups that had sufficient names for analysis: dwarves, elves, hobbits, and men. Although the elf group included names based upon both Sindarian and Quenya, the phonology and syllable structure of the names are so similar that we combined them. Tolkien claimed that Sindarian was influenced by Welsh, Old English, and Old Norse. The language, Quenya, had similar influences, with the addition of Finnish. Our choice to combine them seemed further justified when it became difficult to determine which language Tolkien used to create some elvish names. The same was also true of the men names. We created an "other" category that included orcs, ents, and horses. This group functioned as a control group against which we could compare the other groups. Some groups (e.g. dogs, ravens, goblins) had so few representatives ($n = 14$) that we left them out of the analysis. Therefore, 183 names were used in the final analysis. In Table 1 we list the number of names along with the longest and shortest names in five of Tolkien's name groups.

It is important to review the findings of previous research that has examined phonemes and syllables in Tolkien's characters. Next, we will examine Tolkien's names as whole units using the same IBM name recognition software that is used in airports.

TABLE 1
CHARACTERISTICS OF THE NAMES IN TOLKIEN'S FIVE GROUPS

Name	N	Longest name	Shortest name
Dwarf	23	Faldor (6 phonemes)	Oin (2 phonemes)
Elf	47	Celebrimdor (11)	Osse (3)
Hobbit	18	Bandobras (10)	Bob (3)
Man	74	Bladothin (9)	Tom (3)
Other	21	Bregalad (8)	Arod (4)

Examining phonemes and syllables

Wilcox et al. (2013) found that names made up by a single author are different from those that appear in natural language contexts. This finding is consistent with the work of Fudge and Shockey (1998) who have shown that languages differ greatly in the syllable structures permitted and in the morphophonological positioning of consonants and vowels within syllables.

Wilcox et al. (2013) found that Tolkien's patterns were more limited and predictable than names in authentic contexts. However, one drawback of this study was that all the invented names taken from Tolkien's works were examined as a single group. Cross-linguistic comparisons of the language groups were not performed. Moreover, the previous study did not examine the frequency with which certain phonemes were used within each distinct group of names; nor whether these groups of names differed in terms of syllable structure. For this reason, Smemoe et al. (2014) undertook another study in which they compared the phonotactic patterns and probabilities of the five different language groups of Tolkien. These analyses also included frequency counts of the most common syllable structures.

Results suggested that Tolkien's phonoprint is comprised mostly of these sounds: /b/, /r/, /n/, /l/, /a/, / æ /, / ɛ/; onsets /br/, /gr/, /gl/, and /dr/; and codas /n/, /m/, /l/, /r/, /nt/, /rn/. Although not all of the names he invented followed this pattern, a large number did, despite his conscious effort to invent languages that differed from each other. Such findings suggest that Tolkien was bound by his own natural leanings towards some phonemes over others as he invented these names.

Examining names as whole units

Along with examining phonemes and syllable structures, it is important to also consider names as whole units. When a passenger purchases an airline ticket he or she must provide a full name and birthdate. In most cases, this information is then analyzed by IBM name recognition software that not only checks these data against current no-fly lists posted by governmental agencies, but also determines the probable language background of each name. This is done by comparing the name on the ticket with the 800 million names from all over the world in IBM databases. The software considers multicultural name datasets according to both culture-specific name data and linguistic rules. This approach is based on over 20 years of research and focuses on knowledge-based name matching to improve success rates. It takes into account likely variations, multiple pronunciations, and recognizes written, hand-keyed, and oral interpretations of name data. In addition, it includes transliterations from Arabic, Cyrillic, Greek, Hangul (Korean),

and Kana (Japanese). The software ranks results and provides a score indicating how confident a decision-maker can be that the correct cultural heritage, country of association, and gender have been identified. The current study utilized this software thanks to the support and assistance of Kemp Williams, PhD, a computational linguist who works with global name recognition at IBM, who wished to be acknowledged, but not included as a co-author.

Williams entered the names we provided – authentic names from the nineteenth century census and also character names created by Tolkien. IBM software then identified which languages were implicated by each of the names. Some names were linked with multiple languages. These are referred to as "non-generic" responses. Additionally, the software provided a "generic" category when there was no match for names with any known languages. We examined both categories.

First, we looked at the number of generic responses within the dwarf, elf, hobbit, man, and other groups. We compared these five sources to determine how well the name groups separated from one another. For natural language groups we might expect greater variance in the proportion of unique names in each group. In fiction, where the names are created by the same author, we might expect the name groups to be more similar to one another in the proportion of unique names.

Second, we looked at the languages associated with non-generic responses. All identifications from the software are for modern languages. This becomes problematic when classifying language backgrounds for names that Tolkien claimed were based on ancient languages. Nevertheless, it is interesting to see what languages the character names matched.

Results and discussion

As expected, each word unit was returned with at least one probable ethnic group and language origin. No generic responses surfaced for any of the authentic names that came from nineteenth century census data. However, when Tolkien's names were entered only some were linked with modern cultures and languages. Generic classifications were assigned to 60 out of 183 Tolkien names. The fact that there were no generic responses returned for 100 authentic names and 60 generic responses within Tolkien's fiction suggested that Tolkien was successful in creating multiple unique names.

However, he was not successful in distinguishing among his name groups. The pattern for generic to non-generic responses was approximately the same in each language group, as detailed in Figure 1.

The chi square value for the Tolkien names was 0.850, with four degrees of freedom (n.s.). There was no significant difference in the number of generic responses within the five groups of names invented by Tolkien.

When considering the non-generic names used by Tolkien, possible language connections were identified by the software. This does not mean that these are the languages the names came from, since the software deals with modern languages, not ancient ones. However, the connections that were made are of interest. Some non-generic names had multiple language connections. There was not a one-to-one correspondence between names and languages. Table 2 shows the frequency of possible language sources for Tolkien name groups. Only responses greater than one are included.
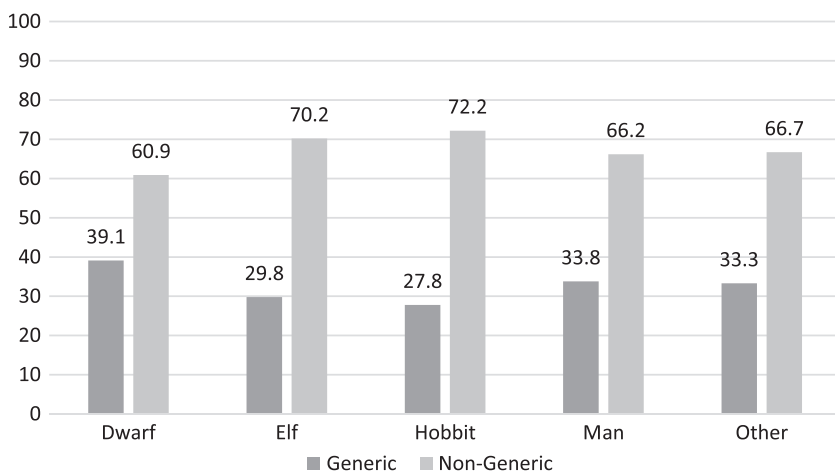
FIGURE 1   Percentage of generic and non-generic names in the five groups

TABLE 2
FREQUENCY OF LANGUAGE SOURCES FOR TOLKIEN'S NAME GROUPS

	Dwarf	Elf	Hobbit	Man	Other	Total
	(*n* = 23)	(*n* = 47)	(*n* = 18)	(*n* = 74)	(*n* = 21)	
Afghani	2					2
Anglo	2	10	4	18	4	38
Arabic	3					3
Farsi	3					3
French		14		12		26
German		10		11		21
Indian			3		3	6
Spanish		14		16		30
Thai			3		3	6

Tolkien reported that names for hobbits and men came from Westron, a language he invented based on Germanic/Gothic, Old English, and Old Norse roots. In this study, the modern languages that surfaced most frequently as possible sources for men were Anglo, French, German, and Spanish. These languages all have a Germanic or a Latin base. This appears to be a good match with Tolkien's claim for names of men. However, these language links are not as strong for hobbits. The most frequent language connections for hobbits were Anglo, Indian, and Thai – the same sources for names in the *other* category. This is not a good match between hobbit names and Tolkien's claims about their origin. It is a good match between other names and Tolkien's claims that names of ents and orks were pidgin forms of his invented languages, and were not meant to sound European. It is interesting that the sources of names for elves were similar to those for men, although Tolkien claimed elf names came from some different language roots (including Welsh and Finnish). For dwarfs, the languages that surfaced more than once were Afgani, Anglo, Arabic, and Farsi. These languages are in line with Tolkien's claim that dwarf names were based on Middle Eastern languages.

Based on examining names with non-generic responses, it appears that Tolkien was able to differentiate among languages, with elf and man names being similar, hobbit and other names being similar, and dwarf names being different from the others.

Conclusion

Personal names are derived from a variety of sources while fictional names in fantasy come primarily from one source, the author. Consciously or subconsciously, it appears that Tolkien used the same phonemes and syllable structures for names regardless of the languages from which they derived. In this way, Tolkien appeared unable to escape his own phonoprint. This pattern continued when we looked at whole word generic responses. But when we examined language responses for whole word units, it appeared that Tolkien was able to differentiate the names into three language groups.

The results of this analysis demonstrate that, although Tolkien may have wished to use sound symbolism (Robinson 2013a) and invented languages to create character names that had the ring of authenticity (Algeo 1982, 2001), he was unable to do so. As primary generator of these names, he was unable to escape his own phonoprint. Even though Tolkien was able to differentiate between name groups, when we considered non-generic responses using name recognition software, he was not able to do so with the precision he claimed.

Tolkien's fictional names do not appear to have the variance seen in personal names that come from a variety of sources. It appears that even a fantasy author as talented as Tolkien, who conscientiously attempted to use "phonetic fitness" to describe characters and create names consistent with the norms of different cultures (Robinson 2013b; Smith 1997), found it difficult to escape his own biases and preferences for some sounds (Smemoe et al. 2014). This suggests that phonoprints merit further research to reveal how these biases are related to an author's perception of what meanings are symbolized by various sounds.

Disclosure statement

No potential conflict of interest was reported by the authors.

Bibliography

Algeo, John. 1982. "Magic Names: Onomastics in the Fantasies of Ursula Le Guin." *Names* 30: 59–67.

Algeo, John. 2001. "A Fancy for the Fantastic: Reflections on Names in Fantasy Literature." *Names* 49: 248–253.

Allan, James. 1978. *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, UK: Bran's Head Books.

Black, Sharon, and Brad Wilcox. 2011. "Sense and Serendipity: Some Ways Fiction Writers Choose Character Names." *Names* 59 (3): 152–163.

Burelbach, Frederick M. 1982. "An Introduction to the Naming in the Literature of Fantasy." *Literary Onomastic Studies* 9: 131–148.

Croft, Janet Brennan. 2009. "Naming the Evil One: Onomastic Strategies in Tolkien and Rowling." *Mythlore* 28: 149–163.

Debus, F. 2002. *Namen in literarischen Werken (Er-) Findung – Form – Funktion.* Stuttgart: Akademie der Wissenschaften und der Literatur, Abhandlungen der Geistes und sozialwissenschaftlichen Klasse, nr. 2.

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: 23–42.

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

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

Jesperson, Otto. 1964. *Language: Its Nature, Development, and Origin*. London: George Allen & Unwin Ltd.

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

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

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

Robinson, Christopher L. 2013a. "Phonetic Metaphor and The Limits of Sound Symbolism." *Names* 61: 189–199.

Robinson, Christopher L. 2013b. "What Makes the Names of Middle-earth So Fitting? Elements of Style in the Namecraft of J. R. R. Tolkien." *Names* 61: 65–74.

Salo, David. 2004. *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.

Smemoe, Wendy Baker, Brad Wilcox, Bruce L. Brown, Paul Hoskisson, and Sky Rodeo Nuttall. 2014. "Naming Practices in J.R.R. Tolkien's Invented Languages." *Journal of Literary Onomastics* 3: 5–23.

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

Tolkien, J. R. R. 1971. "Quenya Phonology." *Parma Eldalamberon* 19: 18.

Tolkien, J. R. R. 2006. *The Monsters and the Critics and Other Essays*. Edited by Christopher Tolkien. London: Harper Collins.

van Dalen-Oskam, Karina. 2013. "Names in Novels: An Experiment in Computational Stylistics." *Literary and Linguistic Computing* 28: 359–370.

Wilcox, Brad, Bruce L. Brown, Wendy Baker Smemoe, Sharon Black, and Justin Bray. 2013. "Identifying Authors by Phonoprints in Their Characters' Names: An Exploratory Study." *Names* 61 (2): 101–121.

Notes on contributors

Brad Wilcox in an associate professor in Religious Education at Brigham Young University. His research interests include elementary reading and writing instruction, education in international settings, and onomastics.

Bruce L. Brown is a professor in the Department of Psychology at Brigham Young University. His research interests cover a wide range of topics within the field of psychology and onomastics. He has co-authored the textbook *Multivariate Analysis for Biobehavioral and Social Sciences: A Graphical Approach*.

Wendy Baker-Smemoe is an associate professor in the Department of English Language and Linguistics at Brigham Young University. Her research interests include first- and second-language acquisition, speech perception and production, psycholinguistics, and onomastics.

Timothy G. Morrison in an associate professor in the Department of Teacher of Education in the David O. McKay School Education at Brigham Young University. His research interests include elementary reading and writing, specifically comprehension assessment and instruction.

 Correspondence to: Brad Wilcox, 316-R JSB, Brigham Young University, Provo, UT 84602, USA. Email: Brad_Wilcox@byu.edu