THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DIVISION OF HUMANITIES, ARTS, AND SOCIAL SCIENCES

THE WANDERERS AND THE ANALYTICS: A LEXICOSTATISTICAL STUDY OF SINDARIN AND QUENYA

CHRISTIAN BRENDEL
SPRING 2014

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Communication Arts and Sciences
with honors in Communication Arts and Sciences

Reviewed and approved* by the following:

Sandy Feinstein
Associate Professor of English
Thesis Supervisor
Honors Advisor

Holly Ryan
Assistant Professor of English
Faculty Reader

Jeanne Marie Rose
Associate Professor of English
Faculty Reader

* Signatures are on file in the Schreyer Honors College

Abstract

Literary analysis of Tolkien's *The Lord of the Rings* and *The Silmarillion* has long incorporated an interest in the usage of language in the texts, including the examination of Tolkien's constructed, fictional Elvish languages. For example, the break-up of language has been seen as a metaphor for the fragmentation of original truth or purity, and the divergence of the various races of Elves is connected to the separation of their languages (Flieger, 2002). In the narrative, the two Elvish languages most used in *The Lord of the Rings* and *The Silmarillion*, Quenya and Sindarin, diverge from a shared ancestor, but evolve in isolation from each other: Quenya develops in the heavenly Undying Lands whereas Sindarin remains in Middle-Earth, a land of relative strife and hardship. Although Quenya is a language heavily influenced by the divine beings who shaped the world, it is Sindarin, curiously, which wields power in Middle-Earth, and is often used to invoke magic. The historical development of these languages has been previously analyzed using the comparative method of historical linguistics, first by Tolkien himself and later by linguists such as Allan (1978a, 1978b, & 1978c) and Salo (2004), who reconstruct the shared ancestor language and examine the relationship between related words or cognates. However, lexicostatistical methods of linguistics—which provide measurements of the rate of lexical change (Swadesh, 1952; Lees, 1953) and the degree to which the vocabularies of two related languages differ (Petroni & Serva, 2011)—have not featured in these analyses of the Elvish languages. Using these methods, this study calculates the rate of decay of the Elvish languages and examines the lexical distance between Quenya, Sindarin, and the root forms from which the words of both languages descend. The results suggest that Elvish decays more slowly than real-life languages, losing only about 6% of core vocabulary per millennium compared to about 20% for natural languages (Lees, 1953). The relatively conservative nature of Elvish can explain Quenya's usage as a language of history. Additionally, the analysis of lexical distance suggests that Sindarin is closer, or purer, in form than Quenya to the hypothetical roots of the ancestor language. Sindarin, therefore, may be more connected to the source of power in Middle-Earth. These results can be used to inform an understanding of the link between the decay of the Elvish languages and the deterioration of the Elvish race, to examine the relationship between language and magic in *The Lord of the Rings*, and to offer new insights into literary analysis of Tolkien's works.

Table of Contents

Acknowledgments

Several faculty, staff, peers, and family members have had a lasting effect on my life. A section that begins to adequately express my gratitude could fill another 70 pages. I would like to take the time to thank those whose assistance most directly enabled me to write this thesis:

My parents and grandparents, for putting books in my hands, no matter how rapidly I went through them.

My mentors at the Smithsonian, Gabriela Pérez Báez and Ives Goddard, as well as the co-directors of the Natural History Research Experiences program, Elizabeth Cottrell and Gene Hunt, and the coordinator, Virginia Power, for giving me the summer of a lifetime.

Leendert Plug, for helping me understand phonological theory, integrating me into the discourse of linguistics, and enabling me to get the most out of my academics abroad.

Neal Woodman, for his shrewd advice in helping me conceive the statistical framework of this thesis and for all of his support before, during, and after my time at the Smithsonian.

Cheryl Nicholas, for being an unerring ally, immeasurable mentor, and ceaseless supporter of both my personal and academic life, and for opening my eyes to the panoptic world.

Holly Ryan, for her unending enthusiasm, encouragement, and dedication, even when I would be a little slow or be a little late, and for helping me to think about it.

Jeanne Marie Rose, for her uninterrupted interest and involvement in my career and life, for showing me the power of rhetoric, and for helping me face the monsters at home and abroad.

Sandy Feinstein, for guiding me on life's journey to *Paradise* and inspiring me to drag on, for completing the quest when defeat looked certain, and for helping me back from Mordor.

Preface

My interest in Tolkien's world grew, as for many other students I know, alongside my interest in linguistics. When I first read *The Fellowship of the Ring* in 2002, I was a kid who was interested in languages: I enjoyed my German course, I devoured the *Teach Yourself Latin* book my grandmother bought for me, and I began noticing curious similarities on my own—why does German *Bruder* look a bit like Latin *frater*?—but it was through my subsequent reading of Tolkien's other works, like *The Silmarillion*, that I discovered that these interests were encapsulated by the labels of philology and linguistics. I was captivated from an early age by the various appendices and essays that Tolkien had written concerning his fictional languages. Every word he constructed was made with intention, and these words fit into the world he was building: the vocabulary of his languages, from the enigmatic Elvish to the malicious Black Speech, was not randomly chosen, but seemed to have evolved organically and according to set rules and constraints. In other words, as the twelve year-old in me might say, they looked and felt like they *belonged*— even then, to me an Elvish word *looked* Elvish, and I knew these words weren't just made up on the spot. Instead, they formed natural connections with the world around them, just as the relation between *Bruder* and *frater* could not be just a coincidence.

This thesis has been informed by other language-based research I've done in the past four years. The common thread of these projects, including this thesis, has been the usage of linguistics, particularly comparative linguistics, as a lens of literary criticism. I've continually visited this topic since I was a first-year student, beginning with a paper on Milton's *Paradise Lost.* In that first college paper, I looked at the etymology behind the name of a specific demon, Tammuz, and how the propagation of this epithet throughout languages over time added to the construction of the demon in books I and II of Milton's work. I again interpreted literature through a linguistic frame when I wrote on *The Knight of the Cart*. I discussed how different translations' usage of pronouns in contrast to specific names led to vastly differing implications about the construction of the titular character Lancelot. Applications of linguistics

in literature have been a constant theme in my work as an undergraduate, and I hope that the pages below are evidence of the value to be gained from using linguistics as a tool of literary analysis.

The methodology I've chosen to use in this project stems from another source. In the summer of 2013, I was a research intern at the Smithsonian National Museum of Natural History, working in the linguistics department with curators Gabriela Pérez Báez and Ives Goddard on Juchitán Zapotec, a native Mexican language in the Oto-Manguean family. My research focused on the methods of adaptation of Spanish loanwords into the genetically-unrelated Juchitán Zapotec language and involved aggregating and systematically analyzing lexical data from a dictionary of the language (to which I also contributed). During this internship, I was inspired to change the focus of my thesis from the traditional comparative method utilized in much historical linguistic research; instead, I decided to borrow inspiration from the more quantitative study I conducted for Juchitán: as data analysis illuminated new facts about Juchitán, I think the lexicostatistical methods used in this thesis offer new, unique interpretations of Tolkien. Additionally, having seen the impact that dictionary development has on the community of speakers and in linguistic research, I intend to utilize the data collected in the creation of this thesis to support the creation of an online, relational dictionary of Eldarin languages. It is my hope that this element will help to generate interest in linguistics for curious readers like me and in exploring both the creative and pragmatic uses of comparative linguistics.

Chapter 1: Introduction

In *The Lord of the Rings* and *The Silmarillion*, J.R.R. Tolkien constructs a fictional world called Middle-Earth and within this world juxtaposes many different languages, both real and fictional. The influence of various nonfictional languages and literatures, such as those of Old Norse, Finnish, and Welsh, on Tolkien's writings has been detailed (Solopova, 2009). In addition to these real-world sources, Tolkien, a linguist, also created his own languages. The construction of these languages was a significant influence on Tolkien's creation of the story of Middle-Earth. In a letter to the Houghton Mifflin publishing company, Tolkien (2000) says that his stories are "fundamentally linguistic" in inspiration; he wrote his stories mostly so that the fictional languages he constructed had an accompanying history (p. 219). As a philologist at Oxford, Tolkien modeled the processes of sound change in his languages after those in real-life languages: his most complete languages, Quenya and Sindarin, which are spoken by the race of Elves, undergo regular sound changes from their ancestor Proto-Eldarin, similar to how Spanish and Italian underwent regular sound changes from Latin. Accordingly, even though Tolkien's languages are fictional, they can be analyzed with real-world methodologies, and the results of such studies can be used to understand his works.

Indeed, Tolkien's languages have long been the subjects of linguistic analysis since the publication of *The Fellowship of the Ring* in 1954, when language enthusiasts first began analyzing the Elvish languages (Hostetter, 2007). The field of Tolkienian linguistics has since broadened and produced materials which describe the structure behind the Elvish languages and the relationship between them (Allan, 1978b & c; Salo, 2004). However, the exclusive usage of the comparative method, which reconstructs hypothetical ancestor languages based on data from attested languages, has until now precluded the application of other methods of historical linguistics and any insights those frameworks could offer. Lexicostatistical methods like glottochronology in particular can provide information on the temporal separation between languages and offer statistical measures of the extent of vocabulary shift

between a set of languages. The values determined from these equations can provide some sense of a qualitative measurement of the genetic closeness between two related languages and suggest the degree of distance between the base vocabulary of these languages, which, as I will later describe, has implications for the social role of these languages in the fiction.

Although the validity of glottochronological analysis, particularly regarding its claim that a subset of fundamental, cross-cultural vocabulary changes at a constant rate, has been challenged (Knut & Vogt, 1962), I think that the methodology is suited for application to the Elvish languages because the framework may be less problematic in a self-contained world, in a closed dataset. Additionally, the concept of language decay, with which glottochronology is concerned, has parallels to the gradual disappearance of a specific race in Tolkien's works. The Elves, by the time of *The Fellowship of the Ring* (and, indeed, after the return of the Noldor to Middle-Earth in *The Silmarillion*) are fading, emigrating from Middle-Earth and maintaining less and less agency as their former power is replaced by that of Men, a younger race[1]. The concept of morpheme decay can be related to this metaphorical decay: is Sindarin, spoken by the Elves who have returned to Middle-Earth, decaying more than Quenya, which is supposedly the purer form of Elvish, having been "closer" to the Undying Land of Aman, where nothing dies or decays? If, by contrast, Quenya is shown to be changing, or decaying, more quickly, what would such a finding mean in terms of the decay of the race, or the "purity" of the origin of this "High Elvish" language? Alternatively, if Sindarin is indeed more decayed than Quenya, why is it consistently used as a language of power—of magical runes and constitutive pronouncements that can literally change the world—instead of its sister language Quenya? Quenya, if it is the closest to the gods who shaped the world, might be expected to wield more power over it; why, then, would Sindarin, if its state of decay were greater, exert such power in the world? Where does the source of the power of language lie in Tolkien's *legendarium*, and does a language being "closer" to it, in some quantitative way, influence the degree of power a speaker of that language wields? In this thesis, I try to answer these literary questions

[1] Indeed, perhaps the Elves never had as much freedom as Men to effect change in the world since they are bound to fate in a way Men are not; see Flieger, 2002, pp. 52-53

through a lexicostatistical analysis of Sindarin and Quenya, using the concepts of glottochronology and lexical distance.

Chapter 2: Background

Grounding Middle-Earth

Of the primary sources for this project, two are the focus of my critical lens: *The Lord of the Rings* and *The Silmarillion.* Of these two, I give particular attention to the latter due to its delineation of the origin of the various Elven races and its explicit notes on language. *The Silmarillion* was made available in 1977, over twenty years after the publication of the final volume of *The Lord of the Rings*, and was compiled by Christopher Tolkien after his father's death. Though *The Silmarillion* was published after *The Lord of the Rings* and is in some ways a work best understood in light of the events that occur after it, the events that it describes happen before those in *Lord of the Rings.* I will therefore outline the relevant plot and key ideas of both of these works which serve as context for my subsequent discussion of the linguistic questions introduced above.

The Silmarillion

The Silmarillion narrates the beginning of Arda, a fictional early Earth, our contemporary knowledge of which has supposedly been lost—it tells the history of our own world eons before we began recording history. *The Silmarillion* begins with a creation story: Arda is sung into being by a choir of angels, the *ainur*, who have been created by Eru, the only true god in Tolkien's writings. The entire fate of the world is outlined by this chorus, and through the disharmonious tune of one singer, the evil fallen angel Melkor, the world is tainted, marred, and doomed to an ultimate death. Language and music, then, permeates and shapes the world from the moment of its creation; in fact, it is literally through words[2] and song that the world is created, and thus the primacy of language and the importance of the act of speech are established early in the mythos.

[2] Or, more specifically, it is through *logos*, according to Flieger, 2002; I will discuss the implications of this concept later in this chapter.

After the creation of the world, many of the *ainur* come to Arda in physical form to keep Melkor, who tries to shape Arda according to his own melody, from corrupting the entire land. The fourteen greatest of these angels, the *valar* (including figures of significance like Varda, also called Elbereth in Sindarin), as well as lesser spirits, called *maiar* (including important characters in *The Lord of the Rings*, like the wizards Gandalf and Saruman, the dark lord Sauron, and the monstrous balrogs), prevent Melkor from wholly destroying the world, but not before he is able to extinguish the first source of light, the Two Lamps, and hide away in a fortress, Utumno, deep in the ground. The *valar* situate themselves in the far west of Arda, in the Undying Lands, the blessed island of Aman, or Valinor, as they repair the world and guard against Melkor in preparation for the coming of the children of Eru: namely, the Elves (and, later, the Men).

Discovered by Oromë, one of the *valar,* on the shores of the lake Cuiviénen, the Elves awaken during this time of darkness (before the restoration of light with the creation of the Two Trees of Valinor, the only light was from the stars). They begin to speak, calling themselves the *quendi,* perhaps meaning 'those who form words with voices' (Tolkien, 1994, p. 391), and spontaneously generate a language unrelated to the Valarin language of the *valar* (which is considerably less attested, or present in the literature, than the Elvish languages).The *valar,* attempting to spare these Elves from the machinations of Melkor, invite them to travel west to Valinor. This action prompts the first of many divisions of the Elves[3]: most choose to follow the *valar*, but others, the *avari* 'unwilling', stay at Cuiviénen. The Elves that travel to Valinor are further divided into three groups each led by a patriarch chosen by the *valar*: the *vanyar* 'the fair [ones]', the *noldor* 'those with knowledge', and the *teleri* 'those who come last'. Of these groups, members of the Teleri linger in Middle-Earth to search for their lost leader, Elwë (in Sindarin later called Thingol, his more usual name in the text), who has fallen in love with a *maia* spirit, while the Vanyar and Noldor proceed onwards to Aman. Ultimately, some of the Teleri decide to stay with Thingol in Beleriand,

[3] This is the first division noted in *The Silmarillion*; in "Cuivienyarna", another version of the awakening written by Tolkien (1994) and reproduced in *The War of the Jewels*, the first three Elves to awaken become leaders, choose companions, and begin to divide the Elves before the Valar even arrive.

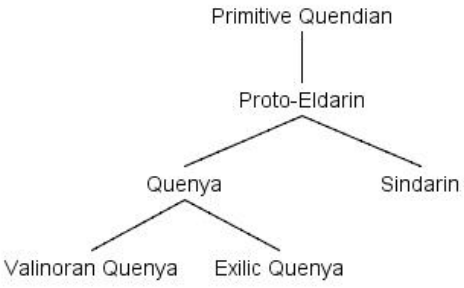
the westmost province of Middle-Earth next to the sea which separates it from Aman. The Vanyar and Noldor go on to Aman and are called the *calaquendi* ‘elves of the light’, whereas the Teleri who remain behind at this last stage of the journey come to be called *sindar* ‘the gray [ones]’.

This division among the Elves is the most central to the use of linguistics to explore Tolkien’s world. At this moment, the remaining Elves who have been journeying to leave Middle-Earth speak a mostly-unattested language which is termed Common Eldarin or Proto-Eldarin. When the *calaquendi* and the Sindar diverge, so too do their languages. This divergence is the origin of Quenya and Sindarin. Quenya is spoken by the Elves in Valinor and is influenced by Valarin, the language of the holy *valar*. Sindarin develops along separate lines and does not borrow Valarin vocabulary (at this stage), nor is its phonology affected by contact with Valarin. A more detailed description of Quenya and Sindarin, including the full phonological inventories and their development from Proto-Eldarin, is given below in the section “The structure of the Eldarin languages and the reconstruction of Proto-Eldarin”. Which language (Quenya or Sindarin) differs more from original form—the basis of this measurement will be qualified in the section “Lexicostatistical methods: glottochronology and lexical distance”—remains to be determined and is the main question that this thesis addresses. Quenya’s social connotation as a prestige language evokes the purity of Valinor, but the extent of its decay from Proto-Eldarin, as compared to its sister language Sindarin, must be investigated in order to determine if linguistic fact correlates with the authority given to it by culture.

Quenya, which is shaped by the Undying Lands, is eventually reintroduced to Middle-Earth. After the awakening of the Elves, the *valar* capture Melkor and imprison him in Aman. Later, over a thousand years after the arrival of the Elves in Valinor, Melkor is freed, feigning repentance for his actions, but his jealousy of the prosperity of the Elves in Valinor leads him to work to corrupt Arda yet again. Fëanor, the son of the king of the Noldor, had crafted the silmarils, brilliant jewels containing the essence of light from the Two Trees of Valinor. Driven by his envy for these jewels, Melkor destroys the Two Trees of light and, after killing Fëanor’s father, steals the silmarils—now the only remaining

container of the essence of light in Arda—and returns to Middle-Earth. Fëanor, who has been influenced by Melkor's corruption, works the Noldor into a frenzy and swears the Oath of Fëanor: he and his sons would reclaim the silmarils and kill anyone—including other Elves and the *valar* themselves—who keep the jewels from the Noldor. This oath leads many of the Noldor to sail back to Middle-Earth, slaying some of their Teleri kin on Valinor who had refused to lend them their ships. As the Noldor in exile return to Middle-Earth, there is another linguistic division—the Quenya of the exiles (Exilic Quenya) diverges from Valinoran Quenya, still spoken by the Elves on Aman (see Figure 1). Unless otherwise noted, Quenya when used in this thesis is synonymous with Exilic Quenya, since it is the variety described in most of *The Silmarillion* and the dialect of Quenya used in the entirety of *The Lord of the Rings.* Figure 1 shows the genetic relationship between these languages:

Figure 1 – Eldarin language family



The exiled Noldor eventually are reunited with the Sindar, still led by King Thingol, who have set up a kingdom in Beleriand in the west of Middle-Earth. Throughout the subsequent ages, Sindarin becomes the *lingua franca* of the Elves in Middle-Earth, spoken by not only the Sindar, but by the Wood-Elves (other Teleri Elves who had abandoned the journey to Aman before the Sindar) to the east of the Misty Mountains, and, eventually, by the Noldor themselves. The adoption of Sindarin as a vernacular by the exiled Noldor is propelled by the linguistic politics of the Sindar, whose policies promote the usage of Sindarin over Quenya: King Thingol in fact outlaws the open usage of Quenya in Beleriand, and the spoken language subsequently falls out of usage (Salo, 2004). Quenya, though, remains in a position of

respect due to its association with the Noldor, who are the most skilled and wise of the Elves in Middle-Earth, and, more importantly, with the Undying Lands to which all Elves will eventually go.

Though the Elves may be immortal, their culture and language nonetheless change, and a linguistic reading of Tolkien informs an understanding of the dynamic status of Elves within Middle-Earth and their capacity to change their world. The original Primitive Quendian divides into Quenya and Sindarin (through several intermediaries, including the last common ancestor Proto-Eldarin). Quenya is spoken by the Elves who journey to Valinor and see the trees of light, whereas Sindarin is spoken by the Elves who remain in Middle Earth and never see the light. Because the speakers of Quenya have seen the light, the social status of their language differs from that of Sindarin, which has never left Middle-Earth, and the structures of the languages differ, as well. Quenya, accordingly, has loanwords from the genetically unrelated language Valarin and becomes a sort of Elvish Latin, respected and preserved as a language of culture and art in later periods of Middle-Earth history, whereas Sindarin becomes the vernacular. The distinction of its speakers having seen the light shapes the social function and composition of each language.

The significance of this light, both in a literal sense and as a metaphor for the original act of creation, has been considered in prior scholarship. Specifically, Flieger (2002) connects the divergence of the Elvish languages to the sundering of the Elven race. In *The Silmarillion*, she proposes, the concept of light—and the literal physical presence of light—is particularly important in the construction of the Elven race. As can be seen from my summary above, light is an object of considerable importance in the early plot of *The Silmarillion,* and the gradual lessening of its purity is a direct result of Melkor's machinations: the quality of the light in the original Two Lamps is diluted first by the Two Trees and finally by the sun as in the real world, following the destruction of each of the earlier sources by Melkor. Light becomes, in other words, tainted and declines from its full, original splendor. The Elven race is fundamentally linked to light when the first created individuals awaken on Middle-Earth: the first word in the first language of the Elves, Primitive Quendian, is *ele* 'behold', spoken in response to the sight of the stars in the sky. The

corresponding root EL is of high importance in the descendent languages. For example, in Sindarin it is seen in *Elbereth*, the name of the aforementioned *vala* Varda, which is a word of power used as a ward against darkness by several characters (including a specific instance I discuss below) in Tolkien's books. The primacy of light helps define the Elven language and their race: indeed, in Quenya, the Elves are called *eldar*, 'people of the stars', from this same root. Flieger (2002) writes that the act of speech in response to the stars "gives the Elves a sense of themselves as perceiving creatures" (p.74), paradoxically "links the Elves with their first inspiration [light]", creates "a distance between the two, and suggests a greater sense of the separation between the Elves and the light" (p. 76). Flieger (2002) suggests that language in Middle-Earth is filtered like this light: as the light dims and as language changes, the original purity and faithfulness to the original form degrades, but just as white light can be splintered into colors, so too does the shift in language allow for more nuanced distinctions to be made.

This initial speech act establishes a significant connection between the Elves, their world, and their language, and begs for further study. The utterance made in response to the natural world both acknowledges the existence of the natural world, tying the Elves' own existence to it, and also delineates these speaking beings as distinct from it. Language distinguishes them into further subdivisions, such as the *calaquendi* and *moriquendi* ('dark elves', chiefly those who were unwilling to begin the journey, but also technically including the Sindar, as well). In Tolkien's works, words act as supernatural performatives, as well: for example, writing gives objects special power, seen in the ability to lock doors supernaturally (the enchanted gate at the Mines of Moria) and, through the corruption of Elvish writing, the creation of the One Ring (the evil, magical object which ultimately must be destroyed in *The Lord of the Rings*). Language gives Elves, and other races, the ability to change and affect their world in ways otherwise impossible, and perhaps as their language changes, so too does their ability to affect—and their tendency to be affected by—their world.

The Lord of the Rings

While much of the primary plot of *The Lord of the Rings*—the quest to destroy the One Ring—does not concern Elvish as explicitly as *The Silmarillion*, several concepts and events in *The Lord of the Rings* illustrate the social role of the Elven languages. Though *The Silmarillion* provides much of the basis for my interpretation of the role of Quenya and Sindarin, *The Lord of the Rings* provides several interesting exigencies that raise questions regarding the usage of Quenya and Sindarin long after Quenya has ceased to have a functional role as the vernacular in any Middle-Earth Elvish society.

In *The Fellowship of the Ring,* the Elves have a very different role compared to their primacy in *The Silmarillion.* The kingdom of Doriath in Beleriand where the Sindar once lived was long ago destroyed, and the exiled Noldor have been flung further apart in their diaspora. While many Elves have accepted their fate—to fade, as a race, from Middle-Earth and allow Men to rule—and have sailed back across the sea to Valinor, others remain. Although some Elves live in other locations like Mirkwood, the two main bastions of Elves seen in *The Fellowship of the Ring* are Rivendell—an idyllic refuge separating the safety of Eriador from the dangers to the east, which is closer to Mordor, the land of Sauron—and the forest of Lothlórien. These are primarily Elven communities which temporarily host Frodo and the fellowship of Men, Dwarves, Elves, and Hobbits (as well as the wizard, Gandalf, a *maia*) that travel with him.

Both of these cities represent Elvendom—they are the last remnants of what was once a powerful force in Middle-Earth. However, the names used to refer to Rivendell and Lothlórien—*imladris* ‘deep valley of the cleft’ and *lothlórien* ‘the dreamflower’, respectively—are, curiously, in Sindarin. Indeed, outside of formal occasions and invocations (such as the lament Galadriel sings in Quenya as the Fellowship leaves Lothlórien), when a form of Elvish is spoken in these areas, it is usually Sindarin, such as when Glorfindel, the elf from Rivendell, greets Aragorn, or when Frodo hails the elf Gildor, which I will discuss below. Such observations prompt sociolinguistic questions that can be used to frame analysis

of Tolkien: for example, if these places preserve the greatness of the Elves, why would they use Sindarin, the common language, over Quenya, the language of the elevated Elves?

One reason could be that Quenya seems to have little power as an instrument of, for lack of a better term, "magic" in Middle-Earth. Specific instances of Sindarin usage challenge the idea that Quenya has tangible perlocutionary power (in this case, the ability to perform "magical" actions) due to its incubation in Valinor. One recurring phrase of significance is *O Elbereth Gilthoniel* (and other variations, generally varying in punctuation, in word order, and in the usage of *A* as the interjection). This invocation, translating to 'O Elbereth Starkindler', is a reference to the goddess of the stars who is called Varda in Quenya. As described above, Varda is a significant figure to the Elves, having created starlight, the first light seen by the Elven race at the time of its awakening. The vocative phrase is first seen in a song sung by a wandering party of Elves led by Gildor, which Frodo translates below:

> Snow-white! Snow-white! O Lady clear!
> O Queen beyond the Western Seas!
> O Light to us that wander here
> Amid the world of woven trees!
>
> Gilthoniel! O Elbereth!
> Clear are thy eyes and bright thy breath!
> Snow-white! Snow-white! We sing to thee
> In a far land beyond the Sea.
>
> O stars that in the Sunless Year
> With shining hand by her were sown,
> In windy fields now bright and clear
> We see your silver blossom blown!
>
> O Elbereth! Gilthoniel!
> We still remember, we who dwell
> In this far land beneath the trees,
> Thy starlight on the Western Seas. (Tolkien, 2005, p. 79)

On meeting Frodo, Gildor goes on to explain that he is a member of the house of Finrod, a Noldor line which had participated in the kinslaying in Valinor and had sworn the Oath of Fëanor. He remarks that, in fact, the entire party are "Exiles" who are tarrying in Middle-Earth before their eventual return to Valinor.

The song itself also locates Elbereth in Valinor, saying that she is "beyond the Western Seas" giving light to those "that wander here" in Middle-Earth: the Elves. Despite the song's concern with historical matters (such as the creation of the stars by Elbereth) and the memory of Valinor, it, nonetheless, invokes the goddess as Sindarin *Elbereth Gilthoniel* instead of the Quenya equivalent *Varda Elentári.* This party of Elves still understands Quenya: when Frodo thanks Gildor for the information the Elf provides, he says "*Elen síla lúmenn' omentielvo*a" 'star shines on the hour of our meeting', a Quenya saying, to which Gildor, amused, remarks "Speak no secrets! Here is a scholar in the Ancient Tongue" (Tolkien, 2005, p. 80). Gildor's lighthearted response provides further evidence for the role of Quenya as an arcane language: it is not a vernacular for these Elves anymore, but a formal language used for phatic communication and, apparently, secret-telling, implying that fewer contemporary people (or, at least, non-Elves) have an understanding of Quenya than of Sindarin. The idea that the usage of Quenya and Sindarin is appropriate or inappropriate in different contexts—Quenya as a language of lore or history and Sindarin as a language of magic or warding, respectively—merits further inquiry.

A potential reason for the stylistic choice of Sindarin *Elbereth Gilthoniel* over Quenya *Varda Elentári* in this song could be to emphasize the geographical distance between the wandering Elves and their former home of Valinor. By singing in a language that had never reached the shores of Valinor, the party of Elves reinforces the idea that they are "Exiles" who dwell in a "far land" separated from Elbereth. The usage of Quenya here, by contrast, would imply a closer connection with Valinor, a relationship which would not reflect the song's concern of separation (both spatially across the sea and temporally through memory). The idea that *Elbereth* is used to imply separation is strengthened by Galadriel, an exile herself, who later comments that "Varda is the name of that Lady whom the Elves in these lands of exile name Elbereth" (Tolkien, 2005, p. 378), suggesting that the usage of *Elbereth* is influenced by the context and environment in which the Elves of Middle-Earth now find themselves.

The choice of the Sindarin invocation, however, is not just an artistic decision: Elbereth's name has power against evil, and in each case is spoken only in the Sindarin form. When the hobbits are hunted

by Sauron's minions, the shade-like Nazgûl (or Ringwraiths) in the valleys around the foot of the watchtower Weathertop, Frodo attempts to ward off the enemy. He cries out "*O Elbereth Gilthoniel!"* and throws himself at one of the Nazgûl. At this, the Nazgûl screams a "shrill cry" and stabs Frodo, but flees with the other assailants. Aragorn (called Strider here) remarks that Frodo's physical attack did not harm the wraith, but that instead "more deadly to him was the name of Elbereth" (Tolkien, 2005, pp. 195-197). However, is it only the goddess's nature as protector of the Elves that fends off evil, or is it also the language in which she is named? In other words, if Frodo had said "*O Varda Elentári*" in Quenya, would the effect (routing the Nazgûl) have been the same?

Other examples illustrate the performative power of Sindarin in contrast to the more formal social role of Quenya. For example, the Mines of Moria, a hidden Dwarven city beneath the Misty Mountains, are magically protected from intrusion by an Elvish riddle in Sindarin engraved on the gates (and with a corresponding answer expected in Sindarin). Later, in *The Return of the King,* Frodo again invokes the name of Elbereth to scare away the spider-monster Shelob. Quenya consistently is constrained mostly to poem and song and typically does not show any pragmatic, magical function in *The Lord of the Rings.* Perhaps due to its authority evocative of the pinnacle of Elven culture, it is instead used in highly formal situations that emphasize the speaker's authority. A prime example of this borrowing of ethos is in Aragorn's coronation, where the newly-crowned king—a Man, not an Elf—recites a speech in Quenya originally given by his ancestor Elendil, the first Númenórean king to return to Middle-Earth. I will consider these examples in further detail after describing the results of my study below.

Linguistic methodology

The methods of historical linguistics can be used to address these questions about Tolkien's fiction. As language is so fundamentally woven into Tolkien's works, studying the structure of and relationship between the languages Sindarin and Quenya with this framework informs an understanding of the role of the Elvish languages and their importance within the text. In other words, scientific inquiry into Tolkien's constructed languages informs a literary understanding of the fictional world within which these languages exist. I will survey the principal methodology of historical linguistics, the comparative method, and the alternative, quantitative methodologies of lexicostatistics, the latter of which I primarily use in this thesis.

The comparative method

The method of analysis used by earlier Tolkien scholars has been the comparative method, the foundational theory of historical linguistics since the formulation of the Neo-Grammarian Hypothesis, namely that language change follows regular, exceptionless rules (Millar, 2007). The comparative method allows for the reconstruction of ancestor languages based on a set of data from genetically-related daughter languages (that is, languages descending from the same ancestor, like Spanish and Italian from Latin). By comparing the features of these languages, the original values of these analogous characteristics can be deduced. For example, a linguist may compare a set of cognate words (words which descend from the same word in the ancestor language) in the Romance languages:

Table 1 – Two Romance words (orthographic representation)

Gloss	Italian	Spanish	French
'iron'	*ferro*	*hierro*	*feu*
'I do'	*faccio*	*hago*	*fais*

These words are genetically related and bear the same general meaning, though the specifics of usage, especially regarding the use of the verb 'to do', vary from language to language. Nevertheless, these

words look very different. For example, the initial sound of each word differs in Spanish compared to Italian and French: where Italian and French have the phoneme /f/, Spanish has Ø (a null value, or silence). The comparative method assumes that sound change is regular and follows consistent rules particular to the development of the studied languages: exceptions to these rules are cause for reformulating the proposed laws. It is preferable, then, to propose a rule that explains this difference rather than to assume that the change is random and spontaneous.

It appears that where Italian and French have /f/, Spanish has nothing. Without the original Latin forms from which these examples derive, it is not immediately clear which reflex (a word or sound in a daughter language) is more conservative (that is, older). Was an /f/ inserted word-initially Italian and French, or did the reconstructed initial **f* become silent in Spanish? When multiple such rules can logically be formulated, the comparative method selects the one that can be stated in a way most consistent with linguists' knowledge of how language systems work and how accurately the generalized rule can capture the facts in the specific languages being studied. We cannot say that vowel-initial words in the proto-language gain an initial /f/ in Italian and French; there are words that begin with a vowel in Spanish and yet have phonemes other than /f/ in Italian and French (for example, the set of Sp *amigo,* It. *amico,* and Fr *ami* 'friend' are all vowel-initial). However, we can say that initial /f/ was deleted in Spanish. This rule would suggest that the original phoneme was **f* in the proto-language, and that **f* became /f/ in French and Italian while it became Ø in Spanish word-initially. In this case, the reconstruction can be verified: hypothetical **f* corresponds to attested Latin /f/, as can be seen from the Latin word from which these forms descend (L *ferrum* 'iron and *facio* 'I do'). By looking at particular phonemes in these words (reflexes of an original, common sound), a hypothetical reconstruction of the original ancestor sound is proposed. As another example, consider the following set in Table 2 – Set of cognate words (the representations are phonetic):

Table 2 – Set of cognate words (phonetic representation) (Millar, 2007, p. 261)

Gloss	**Portuguese**	**Spanish**	**Catalan**	**French**
'court'	*korti*	*korte*	*kor*	*kur*
'high'	*altu*	*alto*	*al*	*o*

If we assume these words to be cognate, we can examine the state of the voiceless plosive /t/ present in some of these forms. We can see this set of correspondences in Table 3:

Table 3 – Set of correspondences of a given phoneme (phonetic) (Millar, 2007, p. 262)

Portuguese	**Spanish**	**Catalan**	**French**
-t-	*-t-*	-∅-	-∅-

Where Portuguese and Spanish have /t/, Catalan and French have nothing. A comparative linguistic analysis would reconstruct **t* for the hypothetical original sound because it is possible to explain the absence of /t/ in Catalan and French as the result of a deletion rule (perhaps **t* is deleted word-finally, or in linguistic shorthand, **t* → ∅ / _#), whereas it is less intuitive to suppose that /t/ was added word-finally in Portuguese and Spanish from nothingness. Such analyses allow us to reconstruct proto-words and, in the case of Proto-Indo-European (PIE), the hypothetical ancestor of many European languages, proto-languages that have never been written down have been reconstructed.

The same methods of comparative analysis have been used to reconstruct Proto-Indo-European. Using sets of phonemes which have been found to correspond in several daughter languages, the phonemic inventory of Proto-Indo-European was reconstructed (and, in the case of several remaining areas of inquiry like the nature of the so-called "pharyngeal" consonants, continues to be described and revised). Take the following set (Table 4):

Table 4 – Reflexes of Indo-European languages and reconstructed PIE plosives (Fortson, 2009, p. 55)

Sanskrit	Avestan	Greek	Latin	Welsh	English	Armenian	Lithuanian	PIE
bh	*b*	*ph*	*f*	*b*	*b*	*b*	*b*	**bh*
dh	*d*	*th*	*f*	*d*	*d*	*d*	*d*	**dh*
h	*z*	*kh*	*h*	*g*	*g*	*j*	*ž*	**ĝh*
gh	*g*	*kh*	*h*	*g*	*g*	*g*	*g*	**gh*
gh	*g*	*ph*	*f*	*g*	*b*	*g*	*g*	**g^{w}h*

The phonemes contained in this table have been found to correspond—that is, they are reflexes of an original proto-sound—in cognates in all of these languages. The set of phonemes in the final column are unattested, proposed original sounds from which the other phonemes developed. By examining data from multiple languages in multiple subfamilies, facts which are otherwise obscured can be discovered. For example, if a linguist were to examine only Armenian and Lithuanian—which themselves are not particularly closely related, one being in the Armenian family and the other in the Balto-Slavic family—he might reconstruct *g for the final two rows: these languages do not maintain any distinction between what we know were two different proto-sounds. However, when the rest of the corpus is examined, other languages are seen maintaining the original distinction: Latin has a reflex for each proto-sound (/h/ and /f/) where Armenian and Lithuanian have only one (/g/). Other languages, like Greek and English, also maintain the distinction. The existence of this distinction, then, leads us to believe that the distinction existed in the proto-language. The importance of using all available data for historical reconstruction cannot be overstated.

Lexicostatistical methods: glottochronology and lexical distance

Glottochronology, a lexicostatistical method which yields quantitative data, involves identifying a set of words based on one commonly-used set of lexical items in genetically related target languages. This

set, often called a Swadesh list after the inventor of the method, consists of words for cross-cultural, universal concepts like ‘hand’ and ‘water’. The number of cognate pairs for given concepts in this list can be used to estimate the chronological separation of the target languages. The decay of morphemes (the smallest particles that can bear meaning, like *–ful* in *wonderful*) is analyzed through lexicostatistical equations and is analogous to radioactive decay in physics. In fact, the analogy to physics is quite literal—the equation for simple radioactive decay is used to calculate morpheme decay (Lees, 1953):

$$N = N_0 e^{-\lambda t}$$

N_0 is the number of morphemes at the start of the analytical period, and N is the number of those same morphemes remaining at t, which is the amount of time elapsed since the divergence of the two sets N and N_0. e is the common mathematical constant 2.718. λ is rate constant of decay that applies across all languages. The value of this constant can be derived from the value of another variable, k, which is equal to $e^{-\lambda}$, and represents the fraction of morphemes retained in a language per millennium, which Lees (1953) places at .8048 ± .0176 per 1,000 years. This value was calculated from the average of the rate constants calculated from the application of glottochronology to various languages. In other words, glottochronological analysis based on this constant assumes that at any given point in time, for any language, about 81% of a language’s basic root morphemes will be retained after the passage of 1,000 years. This equation can be rearranged to $t = -\frac{1}{\lambda} \ln \frac{N}{N_0}$, which allows for the calculation of the number of years separating an earlier morphemic set from a later one. For example, if this study were concerned with only either Quenya or Sindarin and if the Swadesh list for Common Eldarin were known, the time separating the parent language and the daughter language could be calculated.

The equation derived from the above logic, the divergence equation, is most important to this study because it allows the calculation of the time separating two related languages. If two lists can be generated, and if the constant of decay is accurate, the supposition of the divergence equation is that the divergence of two sister languages from a mother language can be dated (Lees, 1953):

$$t = \frac{\log F_s}{2 \log k}$$

In this equation, *k* (as described above) is the retention rate, which Lees places at .8048 ± .0176 per millennium. Swadesh (1955) estimates this rate to be .864 ± .0333 per millennium: Swadesh's greater estimated rate of retention reflects his own test list, which is more stable over time than his earlier 215-word list; Lees (1953) also uses this larger, less stable list and therefore calculates a lower rate of retention. Further discussion of the impact of the word list composition on glottochronological analysis can be seen in my chapter on Methodology.

Related to the concept of glottochronology in its function as a quantitative measure of linguistic change is lexical distance. According to Petroni and Serva (2011), lexical distance is "the minimum number of insertions, deletions, or substitutions of a single character needed to transform one word into the other" (p. 55). In other words, lexical distance (when applied to phonetic, not orthographic, representations of words) offers a measurement of difference between words at a phonetic level. Petroni and Serva (2011) provide an equation to calculate this value, drawing from the comparison of strings of text in computer science:

$$D(\alpha_i, \beta_i) = \frac{D_1(\alpha_i, \beta_i)}{L(\alpha_i, \beta_i)}$$

Here, α_i and β_i are each a word for the same concept *i* in two different languages. $D_1(\alpha_i, \beta_i)$ is the Levenshtein distance, determined from an algorithm which yields the number of character modifications (insertions, deletions, or substitutions) necessary to change α_i into β_I. $L(\alpha_i, \beta_i)$ is the length of the longer of the two words. Dividing by the length of the word puts the Levenshtein distance into perspective: for example, a Levenshtein distance of 3 for a word of six characters would yield a greater lexical distance than a word of twenty characters with an equal Levenshtein distance. For example, the Levenshtein distance of the English word /fæt / 'fat' and the German word /fɛt/ 'fat' would be 1: /æ/ is replaced by /ɛ/, a total of two modifications. The lexical distance, then, would be $\frac{1}{3}$ or 0.33. For another two words,

English /brʌðər/ ‘brother’ and German /brudər/ ‘brother’ would have a Levenshtein distance of 2 (two substitutions), but a lexical distance of $\frac{2}{6}$ or, again, 0.33: despite having more modifications, these words have the same measure of lexical distance and, due to their increased length, are just as similar to each other, on average, as the ‘fat’ pair. The normalizing effect of the word length, according to Petroni and Serva (2011), allows lexicostatisticians to gather “sensible results” (p. 56) from the comparison of words. Ultimately, the concept of lexical distance can be used to establish indices of average lexical distance of common vocabulary (for example, with a Swadesh list) in order to produce a statistical measure of the distance between two languages as a whole. While the glottochronological method can help us deduce the time since the divergence of two languages, or postulate on the projected decay of morphemes given a period of time, lexical distance can provide a synchronic, at-present measure of the difference in the lexicons of two languages from a phonetic perspective.

Lexicostatistical methods like glottochronology, when applied to real languages, can be problematic. Petroni and Silva (2008) argue that a flaw of glottochronology is in its reliance on a “subjective” determination of the cognate status of the words in the Swadesh lists; they state that the results of glottochronological analysis are “biased” towards western European languages (like Greek and Latin) due to the abundance of data on these languages which allows for the comparatively easy identification of cognate pairs, whereas other, less-described languages (i.e., not Indo-European) feature less evidence for the linguist to use when determining if two words in two languages for the same concept are indeed cognates (p. 15). Such subjectivity could, perhaps, introduce error into a glottochronological analysis, though I would argue that this so-called subjectivity is based on the assumed reliability of the comparative method. Regardless, the chief criticism of glottochronology disputes the assumption that language decay is a cross-linguistic constant. As the work of Bergsland and Vogt (1962) shows, some attested languages have rate constants of decay which vary from the proposed universal rate constant significantly. For example, Icelandic, a descendent of Old Norse, is well-known in historical linguistic studies for its extreme conservatism: it has “scarcely replaced any ancestral words at all” since Iceland

was settled (Millar, 2007, p. 461). Icelandic retains almost all of the morphological attributes of Proto-Germanic (Haugen, 1990), and its rate of morpheme decay is also low: as demonstrated by Bergsland and Vogt (1962), Icelandic has a rate of change of about 4% per millennium, compared with Lees' (1953) proposed average of about 20% across all languages. This strikingly low rate of change provides evidence against the idea that morpheme decay is more or less constant across languages, and Bergsland and Vogt (1962) conclude that glottochronology is "unreliable for determining the subgrouping of cognate languages and dialects" (p. 126). Outliers like Icelandic challenge the concept of glottochronology itself: perhaps the idea that the rate of morpheme decay is more or less the same across all languages is a flawed premise.

The status of glottochronology, in the form championed by Swadesh and by Lees, is contested in the real world. However, in a closed and self-contained system—such as in Tolkien's fictional world—where only a few languages are documented and no more are to be discovered, compared to the ever-growing corpus of data on real-world languages, the insights that lexicostatistical methods offer can provide relative measures of linguistic decay and difference that could inform a better understanding of the relation of the fictional Elvish languages.

The structure of the Eldarin languages and the reconstruction of Proto-Eldarin

Much as Proto-Indo-European has been reconstructed based on evidence from attested European languages, amateurs and linguists have used these linguistic methods to reconstruct Proto-Eldarin, the fictional ancestor language of Tolkien's Sindarin and Quenya. The various accounts of these languages have been complicated by the continual introduction of new information: much of the work I describe below, chiefly Allan's (1978b, 1978c), was completed before the publication of many of Tolkien's posthumous works, such as *The Silmarillion* and many of the *History of Middle-Earth* volumes, which contain not only new attested examples of Quenya and Sindarin, but also linguistic outlines on their structures. Much as when a new Indo-European language is discovered, when new data about these fictional languages is introduced, prior theories must be revised. As I will examine below, in recent years

the phonological picture in particular has become clearer, but early work that examined the sound system based on the orthography of *tengwar* script was instrumental in early attempts to document the structure of these two languages. I will focus on the inventory of consonants to illustrate issues and problems in detailing the phonological systems of these languages, but many of the same concerns apply to the vowel inventory, as well.

Quenya

The principle work on Quenya sounds is by Allan (1978b) in the sense that his was not only the first significant scholarly text to detail the language's sounds, but is a "remarkably detailed and accurate linguistic description" of the Elvish, especially given that it was produced before the publication of linguistic materials in *The Silmarillion* (Hostetter, 2007, p. 6). Allan (1978b) describes the linguistic structure and development of the sounds and provides a chart of consonants for both Early Quenya (pre-Exile; Table 5) and Third Age Quenya (which is spoken during the events of *The Fellowship of the Ring*; Table 6). I have updated his representations to International Phonetic Alphabet equivalents:

Table 5 – The sounds of Early Quenya based on the orthography of *tengwar*

	Labial	**Dental/Alveolar**	**Palatal**	**Velar**	**Labialized velar**
Voiceless stop	p	t	t^{y}	k	k^{w}
Nasal	m	n	ɲ	ŋ	$ŋ^{w}$
Nasal + voiced stop	mb	nd	nd^{y}	ŋ g	$ŋg^{w}$
Voiceless spirant	f	θ, ʟ̥, s	ç	x	ʍ
Voiced spirant or approximant	v	ɹ, z	j		w
Other sounds		r rd ɬ l ld	r^{y} l^{y}[4]	 h	

[4] This could perhaps be written as ʎ, the palatal lateral approximant, but, like Allan, I've chosen to represent it as a palatalized consonant instead because of Allan's (1978b) later suggestion that the velarized and palatalized consonants in *tengwar* result from the combination of two distinct phonemes (some consonant + /y/ or /w/) in the language.

Table 6 – The sounds of Third Age Quenya based on the orthography of *tengwar*					
	Labial	**Dental/Alveolar**	**Palatal**	**Velar**	**Labialized velar**
Voiceless stop	p	t	t^y	k	k^w
Nasal	m	n	ɲ	ŋ	
Nasal + voiced stop	mb	nd	nd^y	ŋ g	$ŋg^w$
Voiceless spirant	f	s	ç	x	ʍ
Voiced spirant or approximant	v		j		w
Other sounds		ɬ̥ r rd l ld	r^y l^y	 h	

Between these two tables, we see that Third Age Quenya reduces the number of sounds from that of pre-Exilic Quenya. [θ] disappears entirely, merging with [s], and the combination of the labialized nasal plus velar [$ŋ^w$] also disappears from the orthography, as well. The unvoiced [ɬ] also is no longer present, and each column contains only one type of each sound. No new sounds are introduced.

Allan (1978b) cautions that his tables, while they accurately describe the attested sounds in Quenya, do not depict the phonemes of the language. They are based on how the sounds are written in *tengwar*, the script the Elves themselves use to write Quenya. Each of the sounds listed above, then, are delineated by having different orthographic representations (or, more simply, different letters). However, the correlation between how a writing system represents sounds and how the sounds are organized by a language's phonological system can differ widely. Phonemes are the sound units which are capable of distinguishing words in a language. For example, /t/ and /d/ are separate phonemes in English because they distinguish words like *tip* and *dip*, whereas the sounds [l] and [ɬ], the 'light' and 'dark' L-sounds in

the beginning of *let* and the end of *bull*, respectively, are treated as allophones of the same phoneme, /l/. These two allophones of /l/ are in complementary distribution: one variety appears in one phonological context, whereas the other appears only in another context, and there is no overlap between the two contexts. If the 'dark' [ɬ] were to appear word-initially in *let* ([ɬɛt] instead of [lɛt]), the variant would perhaps sound strange to a native English speaker—though, it may not be noticed at all—but would not affect meaning. However, other languages, like Irish Gaelic, treat [ɬ] and [l] as different phonemes /ɬ/ and /l/, and using one of these sounds in place of the other could create an entirely new, distinguishable word (such as Irish *la* /ɬɑ/ 'day' and *leaigh* /lɑ/ 'melt'). Allan's (1978b) outline, then, shows the sounds that occur in Quenya without comprehensively describing how Quenya organizes them into phonemes which can distinguish meaning in the language.

Nonetheless, Allan (1978b) offers some speculation on the status of a few of these sounds as reflexes of the same phoneme. Allan (1978b) writes [x] and [ç] are "best considered" as allophones of the phoneme /ç/, much like German [x] (e.g. *nacht* [naxt] 'night') and [ç] (*ich* [iç] 'I') are allophones of the phoneme /x/ sharing the same orthographical representation (Wiese, 1996). Other examples of allophony, though, are not so straightforward to resolve, particularly in the system of [b], [d], and [g]. These sounds each appear in restricted sound environments: [b] only follows [m]; [d] only follows [n], [r], and [l]; and [g] only follows [n]. Due to the limited distribution of [b] in particular, Allan (1978b) suggests that it could be subsumed as an allophone of a phoneme /v/, as [v] appears to be in complementary distribution with [b]; that is, the environments in which they appear do not overlap, since [v] is never found following [m] in the corpus of Quenya currently available for analysis. However, Allan (1978b) follows this explanation with an alternative classification suggesting that the three sounds [b], [d], and [g] as allophones of the same phoneme, which he transcribes as [ʠ]. The first method of classification yields three phonemes (or two, if [b] is to be treated as an allophone of /v/), whereas this latter method suggests that there is only one phoneme which is realized as three variants. This significant difference shows the uncertain nature of the phonological system in Quenya.

The phonological status of sounds like [x] and [ç], as well as [b], [d], and [g], does not directly impact the calculations of lexical distance: as allophones, they still represent single segments of differing phonetic value. For example, the lexical distance between Quenya *lamba* ‘tongue’ and Sindarin *lam* ‘tongue’ is 0.4, regardless of whether [b] should be treated as an allophone of /v/, or of /ɣ/, or of any other phoneme. However, some sounds in the Elvish languages present additional problems of representation. Allan (1978b) briefly discusses the classification of units represented with digraphs, suggesting that [ty] and [kw] should “probably” be treated as a combination of separate phonemes: /t/ + /y/ and /k/ + /w/, respectively. Similarly, the segments [ndy] and [ŋdw] likely are clusters composed of the three constituent phonemes. Again, the question of the phonological status of these segments does not affect the Levenshtein distance, but one particular representation indeed poses a problem. Quenya contains a palatal series and a labialized series of consonants (Allan, 1978b), and the transcription of these consonants is represented in one of three ways: for example, the palatalized dental or alveolar nasal can be represented as <ny> (the nasal plus the palatal approximant, IPA [nj]) or as <ɲ> (the palatal nasal, a single grapheme, IPA [ɲ]), or as <n^{y}> (a palatalized dental or alveolar nasal, IPA [n^{j}]). The representation of this sound with either one character or two characters is problematic in at least two ways. First, it is uncertain which of these representations more accurately represents the phonetic reality: is the Quenya *ny* pronounced with the tongue against the palate, as [ɲ] represents (much like Spanish ñ), or does the tongue begin against the alveolar ridge or the teeth, as in [n] (much like English [n]), and then proceed to produce the palatal approximant [j], a combination of two distinct phones (sounds) with different places of articulation? As an example, Quenya *sinya* ‘new’ and Sindarin *sin* ‘new’ have varying measures of lexical distance depending on how these words are represented phonetically. If Quenya *sinya* is [sinja] and Sindarin *sin* is pronounced [sin], lexical distance is 0.40—two deletions occur. However, if Q *sinya* is actually [siɲa], the lexical distance instead becomes 0.50. Although the total number of modifications is the same, the normalization procedure magnifies the result: [siɲa] is one character shorter than [sinja], so these two modifications are weighed more significantly in terms of computing the distance. The two cognates seem to be less closely related with this small alteration in representation.

A few arguments support the view that these units should be treated as combinations of consonant plus semivowel. While Tolkien explicitly states that labialized velars are "a combination" (i.e. [kw], not the single phone [k^w]) (Allan, 1978b), it is preferable to look for evidence provided by the internal processes of Quenya itself. In Quenya, long vowels are shortened when they occur in a closed syllable (one ending with a consonant) (Allan, 1978b). As Allan (1978b) states, the first syllable in the Quenya word *nuquerna* 'reversed' is derived from the root NDÚ with the long vowel [u:]. The vowel is shortened in *nuquerna*, suggesting that the first syllable ends with a consonant, which could only occur if the word were syllabified as [nuk wɛr na], splitting the labialized velar. Were the labialized velar a single co-articulated unit, the phonetic realization of the word should be [nu: k^wɛr na], but since the vowel does not remain long in the attested word, the velar evidently is a combination.

The palatal series does not have such a clear answer. Some similar logic suggests that the palatal series should be treated as a combination of two sounds: for example, the primary stress of *hiruvalye* 'you will find' occurs on the penultimate syllable, a position only possible if the syllable is long. As in *nuquerna*, the syllable can only be long if it is closed, meaning that <ly> must be two sounds, yielding [ˌhi ru ˈval je] (Allan, 1978b). However, in other examples, the palatal series must be treated as single units by the same rules: *máryat* 'her hands' has the long vowel [a:] in the penult, a position that should not be possible unless the syllable were open (lacking a final consonant). The word must be parsed as [ma: rjat] with <ry> treated as a single indivisible sound ([r^j], for example), not as [mar jat]). This evidence directly contradicts the logic outlined above for the treatment of the series as compound. For the purposes of his own dictionary project, Allan (1978b) treats the palatal series in general as combinations of two sounds, but suggests that these units are sometimes treated as single due to etymological reasons: the morphemes comprising *máryat*, for example, are *má* 'hand', *rya* 'her', and *t*, a dual marker. Since the sequence <ry> occurs as part of the lexical item *rya* (not as a combination between a morpheme ending in <r> and a morpheme beginning with <y>), here it is treated as a single unit.

The representation of these series as single sounds or as combinations of multiple sounds, as I showed above with the example of Q *sinya* ‘cloud’ and S *sin*, can significantly affect the analysis of lexical distance that I am performing. For the labialized velars, it appears to be unproblematic to treat all such representations (i.e. <qu>) as combinations ([kw]) uniformly. The best method for treating the palatals is more uncertain. While other Quenya words with cognates in Sindarin feature this series, the Swadesh list used for my analysis only features one: *fanya* ‘cloud’, cognate to S *fain* ‘cloud’. Therefore, for the most part, the problem simply does not surface in the data. When I discuss my findings, the degree of uncertainty regarding *fanya* will be noted.

Sindarin

Even though Sindarin developed in isolation from Quenya and later drove the language of the Noldor out of common use, it nonetheless is constantly defined in relation to its sister language. The word s*indarin* itself is not a native Sindarin word: the Elves of Beleriand, once the *only* Elves in Beleriand after the departure of their kin who would later speak Quenya in Valinor, simply called themselves S *in-Edhil* ‘the Elves’ and their language S *Edhellen* ‘Elvish’. When the Quenya-speaking Noldor return to Middle-Earth, they call the Elves of Beleriand the *Sindar* ‘grey [ones]’ because they are descendants of the Elves who abandoned the journey to see the light of Valinor (Salo, 2004); they are between darkness (never seeking Aman at all) and light (arriving in Aman, like the Noldor had), and this twilight state is referenced in their appellation. When the Noldor return with Quenya, although Sindarin eventually displaces the day-to-day usage of the language of the Noldor, Quenya leaves a mark on Sindarin and literally changes how the language and its people are denoted. In the subsequent years, Sindarin, then, is defined not by its own terms, but in relation to Quenya.

The effect of contact on Sindarin is seen not just in external reference to the language, but is also reflected in its written representation. The Sindar develop runes, the *cirth*, prior to the return of the Noldor (Salo, 2004). However, this native method of writing is later displaced almost entirely by the *tengwar* of Fëanor, used to transcribe Quenya (Salo, 2004), despite the prohibition against the spoken usage of

Quenya in Sindar-ruled Beleriand. While Quenya dies as a spoken language, the writing system invented for it is adapted for use by Sindarin. Written Sindarin, then, cannot help but be defined in reference to Quenya.

Additionally, a number of Quenya words are borrowed into the Sindarin vocabulary. The majority of these loanwords appear to refer to knowledge that the Noldor had received in Aman that was either hitherto irrelevant or entirely unknown to the Sindar (Salo, 2004). For example, when Men appear in Middle-Earth, the Noldor are aware of their existence before the Sindar, and thus the Quenya word *atan* 'man' is borrowed into Sindarin as *adan*, as is the concept of death unique to Men: Q *firya* 'mortal' is adopted as S *fair* (Salo, 2004). The forms of Quenya borrowings into Sindarin were adapted to the constraints of Sindarin phonology: since intervocalic plosives are always voiced in Sindarin, Q *atan* becomes S *adan*. In nomenclature, Sindarin often replaced discrete Quenya elements with their Sindarin cognates. For example, the Quenya name *Telperinquar*, consisting of *telperin* 'silver-like' and *quar* 'fist', becomes Sindarin *Celebrimbor*, with the cognate elements *celebrin* 'silver-like' and *baur* 'fist' (Salo, 2004) (the surface mutation of [n] to [m] occurring from assimilation with the labial [b] and the vowel alteration from [baur] to [bor] is a result of a regular synchronic rule). Occasionally, names were imported with the Quenya elements intact without any attempt at mapping of cognates: S *Enerdhil* comes from Q *Anardil*, composed of Q *anar* 'sun' and Q *dil* 'friend'. Sindarin has a cognate for *anar*: *anor*. However, it seems to have no cognate for Q *-dil* , coming from the Eldarin root –NIL, -NDIL 'friend' (Tolkien, 1987), so both elements are brought into Sindarin with mutations that allow the name to conform to Sindarin phonotactics.

Common Eldarin

Prior to the independent development of Quenya and Sindarin, the eventual speakers of these languages spoke a shared language called Common Eldarin. This language is not the first language that the Elves spoke: it is a descendent of the Primitive Quendian language used by the Elves immediately after the awakening of their race at lake Cuiviénen. As the last common ancestor of Quenya and Sindarin,

it was spoken by the Elves in Beleriand during the final stage of their journey to Valinor. The divergence occurs when the majority of the Elves travel over the sea to Valinor (eventually becoming Quenya speakers) while a group of Elves remain in Beleriand (eventually becoming Sindarin speakers). The importance of Common Eldarin is its status as the last common genetic anchor in the histories of these two languages: the extent of difference between Quenya and Sindarin can be seen in the extent of each of their divergences from Common Eldarin.

The structural status of Common Eldarin is even more inconsistently represented than Quenya or Sindarin. While Quenya and Sindarin are attested in *The Lord of the Rings* and *The Silmarillion*—they are used in speech, songs, and names—Common Eldarin does not feature in the texts, and is only described by Tolkien in manuscripts that were not intended to be published and only later made available by his son. However, various attempts have been made to rediscover or reconstruct the language. I use the terms *rediscover* and *reconstruct* to indicate two different types of scholarly work being performed: rediscovery of the language based on notes published from Tolkien himself and theoretical reconstructions postulated by other scholars.

The question of whether or not to value the author's own propositions more highly than reconstructions proposed by others through valid applications of historical linguistics is a difficult one, especially considering that Tolkien's writings on Common Eldarin in particular are scattered across many notes that were never intended to be published (Tolkien, 1987) and that sometimes were written before Tolkien set important details of his fictional world as seen in *The Lord of the Rings* and *The Silmarillion*. For example, in Tolkien's early notes Sindarin was not called *Sindarin*, but *Noldorin*, a language with an entirely different proposed history than the extant Sindarin, and Christopher Tolkien comments that unravelling this history is "baffling" (Tolkien, 1987, p. 383). In some ways, it might be preferred to treat Tolkien as any other scholar attempting to reconstruct earlier stages of the Elvish languages: when his reconstruction conflicts with that of another linguist, the validity of the claim, not the privileged position of the claimant, must be examined. However, Tolkien occupies the strange position of both linguistic

researcher and *creator* of his languages, and the dearth of available materials on Common Eldarin makes his reconstructions, albeit imperfectly represented, valuable information, and these data must be included in any discussion of Common Eldarin. Thus, when referring to information about the shared language that Tolkien has outlined, I call the language Common Eldarin, indicating that the data more clearly represents how the language was actually spoken; in other cases, I refer to it as Proto-Eldarin, a term more closely linked to the reconstructions linguists posit that may or may not have actually existed. In my usage, this practice is akin to the difference between Vulgar Latin, an attested late stage of spoken Latin, and Proto-Romance, the hypothetical reconstructed form of the last shared ancestor of the Romance languages, which are not synonymous terms. At times (particularly after this chapter), the distinction between Common Eldarin and Proto-Eldarin becomes difficult to maintain, and in these cases I simply refer to the language as Common Eldarin.

In my primary source from Tolkien on Common Eldarin, a chapter in *The Lost Road and Other Writings* entitled the "The Etymologies", Tolkien (1987) outlines Eldarin roots. These items are the elements of vocabulary from which Quenya and Sindarin derive many of their own words. For example, the root NAR becomes Q *nár* 'flame' and S *naur* 'flame'. It is unclear the extent to which this list reflects the forms of the Elvish languages seen in *The Lord of the Rings*, and where possible, I have used Salo's (2004) reconstructions, which are in reference to Classical Sindarin. However, although Tolkien's (1987) "The Etymologies" doesn't suggest specific phonological changes between Common Eldarin and the descendent languages, it is indispensable in providing earlier forms necessary for lexical distance analysis.

For this project, the primary rationale for investigating the phonological system of Proto-Eldarin is so that the cognate status of a given pair of words in Quenya and Sindarin can be determined with some degree of certainty. By identifying the laws of sound change which determine how a given Proto-Eldarin sound develops in Quenya and Sindarin, we can establish whether or not certain words are cognates even if they appear to be very different in surface composition. The determination of this cognate status is necessary for the proposed glottochronological analysis.

In the "Outline of Phonetic Development" of Quenya and subsequently in the revised "Outline of Phonology", Tolkien (2010) provides tables of the original "Quenderin" (Common Eldarin) sounds. These phonemes are the proto-sounds from which Sindarin and Quenya derive their phonological inventories. However, while Tolkien's two accounts of the phonological structure of Common Eldarin mostly overlap, they differ in a few key areas. Table 7 presents Tolkien's (2010) two phonological inventories:

Table 7 – Consonants in Common Eldarin according to Tolkien's "Outline of Phonetic Development" (Tolkien, 2010, p. 30) and "Outline of Phonology" (Tolkien, 2010, p. 69)

Stops	Voiceless	p	t	k
	Aspirated	p^h	t^h	k^h
	Voiced	b	d	g
Continuants	Nasal	m	n	ŋ
	Oral	w	l, r	j, **ʒ**†, w
	Spirant			
	voiceless		s	*h*
	voiced		**z**	

† Boldface indicates sounds listed in "Outline of Phonetic Development" only, and italics indicate sounds found in "Outline of Phonology" only.

Tolkien's (2010) notes on this updated phonological system reconcile the divergences from the earlier representation. He explains that [z], seen as /z/ in his earlier account, developed as an allophone of /s/, which became voiced between two vowels in the descendent languages and was transcribed with a different character in the Feanorian alphabet. The [ʒ] is also described as the result of the weakening of [g]. With this information in mind, the "Outline of Phonetic Development" could perhaps show a relatively later development of Common Eldarin before the divergence of dialects, with the "Outline of Phonology" showing an earlier stage before the development of [z] and [ʒ]: the difference could be viewed as internal reconstruction, the linguistic study of change occurring in what could be broadly termed the same stage of a given language[5]. However, the case of /h/ complicates this linear chronological view of the relationship between the two tables. The appearance of /h/ is explained as the

[5] For example, in the past century, most varieties of Modern English, perhaps beginning with late Old English, have lost /hw/ and merged it with /w/, causing the words *wine* and *whine* to become homophones (Minkova, 2004); this change occurred within a language, and we would likely not consider the English of the 19th century and the English of today to be separate languages like Middle English and Modern English. This change, then, can be called internal

remnant of an earlier series of voiceless spirants which merged to /s/ and /h/. The relative chronology of this development is uncertain; /h/ does not appear in what we have regarded as the later stage of Proto-Eldarin. If the phonology presented in Table 7 truly is this later stage, /h/ should be present since the descendent languages retain reflexes of this proto-sound; it is unlikely and unsupported that /h/ becomes ∅ only to resurface as /h/. In the "Outline of Phonetic Development", Tolkien (2010) writes that Quenya /h/ is derived from /k^h/ and the combination [sk], whereas in the "Outline of Phonology", he says that Q /h/ comes from /k^h/, [sk], Proto-Eldarin /h/, and /p^h/ preceding [u] and [ui]. The two tables, then, do not represent different stages of Common Eldarin, but rather provide two differing accounts of the evolution of Common Eldarin into later languages.

While Tolkien's (2010) tables present his intention of the design of Common Eldarin, they are earlier works that, as evidenced by the discussion of /h/, do not always agree with each other. The phonological inventories of other scholars, reconstructed based on evidence from attested Quenya and Sindarin, may not perfectly describe Tolkien's conception of Common Eldarin, but they nonetheless do present an objective, theoretical account of the Proto-Eldarin from which the later languages must have descended. The earliest independent reconstruction of Proto-Eldarin's phoneme inventory comes from Allan (1978a) and is accompanied by the reflexes that the proto-sounds give in Quenya and Sindarin; I will present a modified version of this table of consonants after the discussion below. In his account of the development and structure of Sindarin (Table 8), Salo (2004) proposes another phonological inventory of Proto-Eldarin, which loosely agrees with Allan's (1978a) account (with an important caveat I describe below):

Table 8 – Consonantal phonology of Common Eldarin (consonant combinations excluded) (Salo, 2004, p. 35)

	Bilabials	Alveolars	Palatalized alveolars	Palatalized velars	Velars	Labialized velars	Glottal
Unaspirated stops	p b	t d	t^j d^j	k^j g^j	k g	k^w g^w	

Aspirated stops	p^h	t^h		kh^j	k^h		
Nasalized stops	mb	nd		ng^j	ng	$ŋ^w g^w$	
Nasals	m	n	n^j				
Trills		r					
Fricatives		z					h
Approximants		l		j		w	

Salo's (2004) system (Table 8) differs from Tolkien's later outline (Table 7) most clearly in the presence of /z/, the palatalized series, and the nasals. Salo (2004) details the changes within Common Eldarin that lead to the phonological system he presents, such as the voicing of /s/ described by Tolkien (2010) above. While Tolkien's (2010) accounts both represent earlier versions of Common Eldarin, Salo's (2004) reconstruction is the closest available to the point of divergence of Common Eldarin into Sindarin and Quenya.

The question of the spirantization of the aspirated series ([p^h t^h k^h] > [ɸ, θ, x]) is the primary point that distinguishes Salo's (2004) account from the one proposed by Allan (1978a). Allan (1978a) reconstructs only the fricative series with no mention of their evolution from a prior aspirated series. Salo (2004) places this change as occurring after the divergence of Sindarin from Common Eldarin; in fact, he proposes that it occurred between the Old Sindarin and Middle Sindarin stage. In both of his outlines, Tolkien (2010) writes that the aspirated series was likely completely opened to the respective spirantized forms early in the history of Quenya of the First Age. Therefore, the same development of the aspirates into fricatives occurred *independently* in isolated Sindarin and Quenya after they diverged from Common Eldarin. Therefore, Allan's (1978a) account of the changes from Common Eldarin that produced Sindarin and Quenya phonemes, while a faithful reconstruction based on the materials available to him at the time of his analysis, is complicated by the publication of the later materials by Tolkien and Salo. It is

inaccurate in regard to its representation of the phonological system of Common Eldarin, but is still useful as an account of the correspondences between Common Eldarin to Quenya and Sindarin if we replace Allan's usage of the spirant series with the aspirated series. Allan's (1978a) table (Table 10) is given in the Appendix in full.

Chapter 3: Methodology

In his early glottochronological analysis of rate constant of decay for English, Swadesh used a list of 215 words referring to the same 215 concepts in each language in his study. These words were the most common colloquial terms for a given idea (Lees, 1953). For example, while English might have several words referring to 'dog', like *hound, pooch, cur* and *dog*, all of which are used in modern English, the popular usage of *dog* would lead us to select this particular word for such a list (even though *hound* is cognate with several closely related languages, as in German *Hund* 'dog'). Though Swadesh himself acknowledged the difficulty in drafting a list of words for truly cross-cultural, shared concepts (Sjoberg, Sjoberg, Swadesh, & Sreekantaiya, 1956), the list nonetheless proposes concepts that are assumed to be common to cultures and languages across the world, words for concepts such as 'I' or 'water'.

The problem of dealing with constructed languages like Sindarin and Quenya is primarily in the limited vocabulary available to scholars. Tolkien simply did not construct a lexicon that approaches the size of natural languages. Some words in Swadesh's list which are easy to find in natural languages, such as the word for 'louse', are not present in any Elvish corpus. Fortunately for this study, Swadesh lists with a lower number of words have been employed by linguists since Swadesh's original study. The first of these, the 100-word list, was proposed by Swadesh himself by narrowing his 215-word list to the most noncultural words (Sjoberg et al, 1956). A drastically smaller list, consisting of only 35 words, was also used by Yakhontov who contended it featured the most stable words in Swadesh's list (Starostin, 2010). However, even using Swadesh's narrowed list, I found that 18% of the required words still are not attested in both the Sindarin and the Quenya dictionaries. The omission of such a percentage of words from the proscribed list could affect the estimate of divergence by several hundred years and call into question the efficacy of the study.

Compounding the problem of the lack of availability of these important diagnostic words is the question of whether the available lexical items are the most appropriate words to use in a glottochronological study. Some concepts have multiple related words in a language, and, as prescribed by Swadesh (1952), the most commonly-used word for a given concept should be used for analysis. However, it is difficult to determine whether or not a Sindarin or Quenya word is the most commonly-used word for an idea. Though many of the first-entry dictionary items are not cognates, there are often other words which are clearly cognate. For example, for the concept 'tooth', Sindarin has *anc* and Quenya has *nelet* as first entries, and these are presumably the most common words and are not cognates. However, later in each entry, S *carag* and Q *carca* are listed, and these words are doubtlessly cognate (it is a general finding that where Sindarin has final *–ag*, Quenya has *–ca)*. For glottochronological analysis, then, should the Sindarin and Quenya words for 'tooth' be treated as non-cognate (S *anc* and Q *nelet*), or should the fact that cognates exist later in the dictionary entries (S *carag* and Q *carca*) suggest that the vocabulary item has survived and should be regarded as evidence that the original Common Eldarin word has survived in each of the daughter languages? This question is of fundamental importance in calculating the results of the glottochronological analysis. The number of cognate reflexes in Quenya and Sindarin for a given vocabulary item significantly affects the predictions of the glottochronological equation because it changes the value of F_s in the divergence equation ($t = \frac{\log F_s}{2 \log k}$), which is the only independent variable in the equation since we are accepting *k* as a constant. The value of *t* varies according to the value of F_s. Establishing that not only are two words cognate, but are also *the most commonly used words* for a given concept in each language, entirely affects the estimation of the length of time since divergence.

There is no clear answer for these questions, especially for constructed languages. In performing my glottochronological analysis of Quenya and Sindarin, I used a list that, by necessity, falls between the word counts of Swadesh's revised 100-word list and Yakhontov's 35-word list. Accordingly, for the value of λ, the rate of vocabulary retention, I will use the value Swadesh (1955) calculated for his 100-word list: .864 ± .0333 per millennium. Regarding the question of word popularity, I have interpreted the

first entry of the dictionary form of a word to be the most popular usage, and, therefore, the one most suited for comparison to the related word in the other language. I have not been able to determine, however, if the order in which Tolkien placed translations in his notes is meant to indicate commonness, and, thus, my treatment of the issue so strictly could be faulted because it is rather mechanical and does not take into account any usage notes external to the dictionary sources by other scholars or Tolkien himself. Glottochronology has traditionally attempted to lessen the degree of subjectivity in historical linguistic analysis and, although many of the entries for a given concept do have some cognate listed *somewhere*, I have endeavored to avoid picking and choosing words that seem most appropriate to me. While identifying any cognate as eligible for analysis is tempting, the most common modern English word for 'dog' is not *hound*, and also, I think, the most common Sindarin word for 'tooth' is not *carag*; consequently, neither of these words are suitable for inclusion in a Swadesh list. I realize the very attempt to avoid subjectivity is in itself a subjective decision. A future study could examine the history of each Elvish word and its concordance in the texts to determine a more accurate account of its colloquiality, but I will not include such special treatments in my statistics here (though I will comment, when necessary, on particularly notable cases).

In assembling the list, I used the information in Tolkien's (1987) "The Etymologies", which provides Common Eldarin roots as well as reflexes in Quenya and Noldorin (i.e. Sindarin). I also used the dictionaries provided in Allan's (1978b) *An Introduction to Elvish* for Quenya and Salo's (2004) *A Gateway to Sindarin* for Sindarin, as well as Fauskanger's (2008) English to Quenya wordlist. In the cases where there were discrepancies, I selected the versions provided by Fauskanger (2008) and Salo (2004), since they are the most recent and are based on the most complete linguistic picture of Elvish that we currently have. I entered phonetic representations of these words into Excel, where I determined their cognate status (using the earlier equivalencies suggested by Allan, 1978a, as well as more complicated rules of sound change in Salo, 2004). This list was then fed into a program I developed in Perl (a programming language useful for text manipulation), where the number of alterations between the

phonetic representations of the words was used to calculate the Levenshtein distance between the words in each sister language, as well as between the given Common Eldarin root, if available, and the daughter reflexes. Entries in the Swadesh list that did not contain both a Sindarin and Quenya translation were discarded from the analysis altogether.

Chapter 4: Results

Ultimately, 82 words populated the original list; 1 was eliminated due to insufficient data[6]; 52 were found to be cognate; and 29 were found to be non-cognate. From the ideal 100-word list, 19% of vocabulary items are missing or not included in the analysis. The entire dataset, seen in Table 11, was used for the divergence equation, seen in the subsequent section. The cognate words[7] in this list were used for the analysis of lexical distance, the results of which can be seen in Table 13.

Results of glottochronological analysis

Using the divergence equation, $t = \frac{\log F_s}{2 \log k}$, the time since the divergence of Quenya and Sindarin can now be calculated:

Table 9 – Divergence time of Sindarin and Quenya

	F_s	t (in thousands of years)
Set 1	$\frac{52}{81} \approx$ **.642**	**1.59 ± .43**
Set 2	$\frac{52}{100} = .520$	2.34 ± .63
Set 3	$\frac{71}{100} = .710$	1.22 ± .33

The three sets calculated account for the error introduced by the lack of 19 words from Swadesh's list. Set 1 uses a value of F_s calculated by disregarding the words missing from the corpus. Set 2 proposes a maximum value for *t* by assuming that the 19 undefined words are all non-cognate, a lower rate of retention. In contrast, Set 3 supposes that the languages feature a higher rate of vocabulary retention,

[6] The Swadesh list entry for 'not' has been discarded from this analysis. Quenya and Sindarin feature several words for 'not', such as Q *lá* and S *ú*. In both languages, though, the meaning of Swadesh's 'not' diverges in Elvish into two forms: one for adverbial negation (e.g. Q *lá*) and one as a prefixed adjectival form that negates a quality (e.g. Q *ú*), much like *un-* in English *unnecessary*. Some of these words are cognate between Quenya and Sindarin (Quenya also has *ú*), and some are not, and the two types of 'not' are very different in Elvish languages. It is unclear which form would be best mapped to Swadesh's 'not'.

[7] A few cognates were excluded because I could not find a definite Eldarin root for them: 'I', 'who', and 'what'.

treating the 19 missing vocabulary items as all cognates which have been conserved over time. In each case, *t* is stated with standard deviation to account for the error in Swadesh's (1955) *k*. Set 1, since it does not make any claims about the cognate status of the missing 19 words, should be the most accurate figure. Therefore, the glottochronological analysis suggests that about 1,590 years passed since the break-up of Common Eldarin with an uncertainty of about 430 years.

The results of the glottochronological analysis—the supposition that about 1,590 years have passed since Sindarin and Quenya began developing as separate languages—does not seem to agree with Tolkien's chronology of events, and the high margin of error makes the results less meaningful for dating the divergence. As has been detailed above, the point of divergence of Sindarin and Quenya is the moment the Elves in Beleriand split: some remain in Beleriand, the Sindar, and others travel to Valinor, the Vanyar and Noldor who would go on to speak Quenya. According to Tolkien's (1994) *The War of the Jewels*, the time that passes between the separation of the Noldor and Sindar and their reunion is "well nigh as" 3,500 years (p. 24). This figure does not agree with the glottochronological analysis; the actual passage of time is more than double the timespan estimated from the divergence equation. In fact, the problem is complicated even further because the Sindarin represented in the writings and books examined in this study is Classical Sindarin, a form of the language which developed in the Second Age (Salo, 2004), which begins almost 600 solar years after the return of the Noldor. This dating would suggest that the actual duration of separation is somewhere around 4,000 solar years in total, and even *this* estimate is complicated by the fact that the Quenya available to us is not necessarily the vernacular spoken by the Noldor who returned to Beleriand (although it could be presumed that Quenya did not change significantly between the Noldor's departure from Aman and their arrival to Beleriand).

This complication could be reconciled in several ways. Firstly, we could claim that the "real" account of 3,500 or 4,000 years is simply wrong: perhaps the Elves, or Tolkien, did not document the history accurately. However, much more likely is the idea that the glottochronological constant used for my equations simply is not suited for calculating the decay of core Elvish vocabulary. Indeed, if the

glottochronological constant Swadesh proposed were applied to a highly conservative language like Icelandic, with its low rate of decay as earlier described, the estimated time of separation between Old Norse and Icelandic might be placed at a few hundred years rather than the historically-attested value of over a thousand. Some languages simply change more slowly, due to isolation or other factors, than the languages surveyed by Swadesh. Perhaps Sindarin and Quenya are one of these slowly-changing languages: the proposed rate of decay is too high.

Since we have an approximate value for *t*, 4,000, and a value for F_s, .642, we could calculate a value of *k*, the rate of retention which Swadesh (1955) placed at .864 ± .0333 per millennium for the languages in his analysis, that estimates the rate of decay for the Elvish languages. Using this data, we arrive at $k = .946$, meaning that about six words of core vocabulary are lost per millennium. The value of *t* gained from using this constant instead of Swadesh's approximates the historically attested number of 4,000 years: the new constant estimates that the separation of Quenya and Sindarin occurred 3,990 years in the past. The significance of this newly-calculated constant for literary analysis will be explored in my final chapter.

Results of analysis of lexical distance

I ran two analyses of the Elvish vocabulary utilizing lexical distance. First, I compared Quenya and Sindarin cognates directly to determine which words were the most similar. These results can be seen in Table 12. Future research could use this data to speculate on the relative importance of certain words to Elvish culture within the context of Tolkien's fiction: for example, does the lack of distance between the Sindarin and Quenya words for 'eye' (S *hen*, Q *hen*) imply that the eye is particularly significant to the Elves? This information could be useful for a different type of literary analysis than what I am conducting here. Instead, in addressing questions relating to decay of language and in determining which Elvish language is more resistant to change, I am far more interested in the results of my second analysis, the comparison between Sindarin and Quenya and their ancestor Common Eldarin, because it reveals new ways of understanding the relationship between the three languages. Table 13 shows the lexical distance

between words in each daughter language and the Eldarin roots. The question of which language is ultimately closer to its ancestor, though, is not entirely straightforward. Quenya words are, on average, lexically closer than Sindarin to the Eldarin roots: when considering the number of modifications necessary to produce a Quenya or Sindarin word from an Eldarin root, the average lexical distance of the Quenya list is 0.456 while Sindarin has an average lexical distance of 0.466. In isolation, these numbers are meaningless, but when viewed comparatively, the figures suggest that Quenya, not Sindarin, is closer to Common Eldarin. A different implication is found when we examine the absolute number of words, rather than the average degree of modification, that are most similar to the Eldarin roots. Disregarding the *extent* of distance (i.e. the normalized number of modifications per word which gives us the average lexical distance statistics above), 21 Sindarin words are closer than their Quenya counterparts, whereas only 14 Quenya words are closest.

In other words, more Sindarin words than Quenya words are closer to Eldarin roots. This fact could imply that Sindarin, *not* Quenya, is, then, lexically closer to Common Eldarin. Still, Quenya appears to have more in common than Sindarin with Common Eldarin *words*, not the Eldarin roots used in my analysis, that were spoken when Common Eldarin was a living language. A future study could examine this assumption in more detail, but anecdotally, I have found that when examining reconstructed Common Eldarin words, the full forms which were in use before Quenya and Sindarin diverged, Quenya seems lexically closer. These Eldarin words are sparsely given in Salo (2004) and Tolkien's various writings, but Quenya seems more likely to retain the word endings that often are placed on Eldarin roots to form Common Eldarin words. While Quenya is more faithful in this regard, Sindarin, though, does not feature much of this inflectional baggage: Sindarin tends to streamline its forms, resulting in a vocabulary which resembles the bare Eldarin roots that were not necessarily independent lexical items. For example, for the concept 'fish', the Common Eldarin word is *lingwi*, coming from the Eldarin root LIW 'fish'. Quenya and Sindarin both retain the reflex: *lingwe* and *lim* respectively. Q *lingwe* and CE *lingwi* are almost identical, but the Sindarin form *lim* has much greater lexical distance from the CE word. However,

S *lim* is extremely close to the Eldarin root LIW and does not feature the extra syllable at the end of the word. Similarly, Quenya *elen* ‘star’ appears to match perfectly the Eldarin root ELEN—which I used for this analysis—with the Sindarin *él* ‘star’ differing more significantly. The Eldarin root ELEN, though, itself has a more primitive root EL, also meaning ‘star’ (Salo, 2004), so the Sindarin word again shows a pairing down to the most essential morphemes. Incidentally, treating EL rather than ELEN as the root in my analysis causes Sindarin to become, on average, lexically closer to the Eldarin roots than Quenya, but, according to the principles of my methodology set out above, I’ve utilized ELEN because it appears to be the most commonly used root. Even without more closely examining my choice of roots to identify the most primitive, morphologically-dense roots, the similarity of Sindarin to Common Eldarin can indeed provide insight into Tolkien’s works. In the last chapter, I discuss the implications of these findings as they relate to the predominance of Sindarin as a magical language in contrast to Quenya.

Chapter 5: Implications

The idea that Quenya and Sindarin are slowly changing—and that Sindarin may be "closer" to the roots of language—has implications for the interpretation of Tolkien. The glottochronological analysis and the study of lexical distance each can inform different yet complementary ways of understanding *The Lord of the Rings.* In this last chapter, I concentrate on potential directions in which future studies could apply my results. The first set of results from the glottochronological analysis allows us to compare the Elvish languages as a family to human languages, but it does not give us any new insights on the social role of Quenya and Sindarin individually or on the almost exclusive usage of Sindarin to perform magical acts in *The Lord of the Rings*. Fortunately, the analysis of lexical distance allows us to begin comparing the languages to each other and to Common Eldarin, and both methods provide unique insights into language in Tolkien's works.

Firstly, in considering the Eldarin languages as a group in relation to other language families in Middle-Earth, the proposed glottochronological constant for Quenya and Sindarin (94.6% retention of core vocabulary per millennium) allows us to distinguish the Elvish languages as a group from the languages of Men in Middle-Earth. The increasing usage of languages of Men rather than Elvish languages is a microcosm of the relationship between the speakers' races: the Elves' apex as Middle-Earth's hegemon and dominant culture is behind them, and they are gradually leaving the lands of hobbits and Men, traveling to the blessed isle of Aman in the West. The Elves are fading from Middle-Earth. Unlike the Elvish race, however, the Elvish languages are resistant to decay. The Sindarin in Beleriand as described in *The Silmarillion* remains the chief common language among the Elves, but its position of *lingua franca* is subsumed by the languages of Men: Adûnaic and its descendent Westron (which is rendered as plain English in Tolkien's works), the latter of which serves as a common tongue among humans, Elves, dwarves, and the other races of Middle-Earth (Solopova, 2009). For humans, the increasingly popular usage of Westron relegates Elvish—specifically, Quenya—to special functions,

where it is used in "lore and ceremony" as it functioned during the pinnacle of the civilization of Númenor, the ruined sea-kingdom of ancient Men (Noel, 1974). For example, when Aragorn, the king of the human land of Gondor, is crowned in *Return of the King*, he recites the words of his distant ancestor, Elendil, who fled from Númenor to the shores of Middle-Earth. He sings in Quenya:

Et Eärello	Out of the Great Sea
Endorenna utúlien.	To Middle-Earth I am come.
Sinome maruvan	In this place I will abide,
ar Hildinyar tenn' Ambar-Metta!	And my heirs, unto the ending of the world.
	(Tolkien, 2005, p. 967)

The usage of Quenya in this oath is significant in two respects. The words proclaim the intent of the Númenóreans to "abide" in Middle-Earth until the end of the world. However, the Elves, as mentioned above, are *leaving* Middle-Earth. Using the language of this very group of people to convey a sense of remanence even as they flee is seemingly contradictory. The usage of Quenya in particular is notable: Quenya, unlike Sindarin, developed outside of Middle-Earth. In many ways, although this recitation is given by human kings about the history of Men, the words of Elendil are applicable to Quenya's development, as well. In fact, the oath is, perhaps, a more accurate depiction of the history of Quenya than the story of Elendil's people. The phrase "Out of the Great Sea / To Middle-Earth I am come" implies not a return, but a first arrival. For Men, the arrival to Middle-Earth *is* a return: Men were first created in Middle-Earth. Similarly, Aragorn, when he returns to take the throne of Gondor after a thousand years without a king, is retaking what was lost, not arriving to a new land. The Oath, then, is not entirely accurate for Men—it is, however, an accurate account of a personified Quenya. Quenya developed "out of the Great Sea", far away from Middle-Earth in Valinor. With the Noldor, it "comes" to Middle-Earth. Its longevity allows it to "abide" until the ending of the world, and its "heirs" are, perhaps, the traces of influence it has on the language of Men and their society. Even though the kings are promising the survival of human kingdoms, they are using *Elvish* words to make the oath: Elendil spoke Adûnaic and Aragorn spoke Westron, but neither chooses the language of Men to cement the establishment of a kingdom of Men, electing instead to use the Elvish. With the evidence of the slow

decay of the Elvish languages, we can see now that the decision to use Quenya is not just because Quenya is more prestigious than human languages: Quenya is more resistant to change than the languages of Men, and its usage here allows its message to "abide". In invoking Quenya, both Elendil and Aragorn are associating their kingdoms with an authority whose source of power (or, more literally, point of origin) is outside Middle-Earth, anchoring their reigns to something that has permanence.

Perhaps a rationale for the usage of Quenya in courtly functions is that it simply decays more slowly than Westron, and, like building a monument, a stronger, more hardy material that can withstand the ages is preferred. For this argument to be viable, some assumptions must be made: Elvish must be more resistant to change than human languages. The linguistic picture of Adûnaic, the ancient language of Men, is fragmented because the language is not well-attested; Tolkien simply didn't develop the language as much as he did his Elvish languages. Additionally, information on Westron, the language which is translated into English in *The Lord of the Rings*, is extremely sparse. Ideally, the languages of Men could undergo a similar glottochronological analysis and the resulting rate constant could be compared to the proposed value I found for the Elvish languages. This comparison could provide insight on the relative stability of Elvish versus human languages: if Elvish decays more slowly than the languages of Men, its usage as a ceremonial language—a way of communicating information that must be permanently remembered—makes sense. However, such an analysis simply is not possible without more information on Adûnaic. I make the assumption that the glottochronological constant for the languages of Men in Tolkien is closer to the estimations given by Swadesh (1955) and Lees (1953) since these humans are cast as the ancestors of modern-day, real-world humans—perhaps their languages, too, are more similar to ours than to the Elves'. Still, while it is valid to suggest that the glottochronological constant was higher (that the languages were more conservative) in the time period Tolkien describes for both Elvish and human languages, Tolkien (1994) himself remarks that the Elvish languages changed more slowly than human languages, saying that Quenya had changed in 3,500 years as much as a human language would

have changed in 500 (p. 24), and the slower rate of change for Elvish suggested by my study supports his claim.

While glottochronology helps explain the dialectic between human and Elvish languages, the results of the lexical distance analysis could allow us to understand why Sindarin has more power than its fellow Elvish language, Quenya. Earlier, I suggested the possibility that the usage of Sindarin rather than Quenya in the invocation of Elbereth, the patron goddess of the Elves, stems from a desire of the Elves of Middle-Earth to acknowledge their separation from Valinor: I proposed that through their usage of Sindarin, which never left Middle-Earth, the Elves reaffirm their ties to Middle-Earth even as many of them prepare to sail across the sea to their ultimate destination of Valinor, leaving Middle-Earth behind. While this explanation could help explain the choices the Elves make in which language they use, it does not address the other point I raised: if the power to fight evil comes from holy Valinor, why is the vulgar Sindarin, not the assumedly-purer Quenya, effective in combatting dark enemies like the spider Shelob, and why does it have magical power as the language which is able to open the Gates of Moria?

The measurement of lexical distance suggests a response to this question that could be developed in future papers. As I described above, Sindarin represents a simplification or discarding of ostentation: words are returned to their literal roots (i.e. to Eldarin), and superfluous material is thrown away. Perhaps these linguistic roots are a source of magical power in Arda, and Sindarin's disposal of the extraneous linguistic material that dilutes this power allows it to be a more potent language. This view of Sindarin as the purer Elvish language differs from earlier criticism. For example, as I introduced in Chapter 2, Flieger (2002) argues that literal light in Middle-Earth fragments and dims: it is first provided in its purest form by lamps, then by trees, and finally by the sun. Each source of light is "dimmer than the one before it", and the Elves and Men who live under the light are subjected to "a sense of loss, of estrangement, and ever-widening distance from the light" (p. 60). The splintering of light is not necessarily a bad occurrence: it adds "color and variety to the world" (p. 70), and Flieger (2002) suggests that the fragmentation of the Elvish languages is linked to this splintering. She goes on to characterize the distinction between Quenya

and Sindarin as an opposition between the light of Aman and the darkness of Middle Earth. The Sindar, she says, occupy "the middle ground … between enlightenment and ignorance" (p. 92), since they participated in the quest to reach Valinor—to reach the light—but did not complete the journey, whereas the speakers of Quenya are closer to the light because Quenya developed in Aman. Flieger (2002) concludes that due to this history, as well as the phonological weakening (lenition) in Sindarin of plosives to fricatives, "Sindarin is farther from the light but closer to the activities and concerns of Middle-Earth" (p. 94). However, perhaps Sindarin's physical distance from Aman is less important, in terms of its capacity for magical power, than its relative lexical closeness to an older, less evolved form of Elvish.

The idea that Sindarin is closer to Common Eldarin does not contradict Flieger's (2002) analysis, but complements it. If the fragmentation of light is analogous to the fragmentation of Elvish, the older Common Eldarin roots could be seen as purer. Whereas Quenya adds more verbal baggage to these roots, Sindarin's lack of this extra material can be seen as a clearing away of brush and bramble which obscure the sky, allowing the light to shine through more completely. Sindarin is lexically farther from Common Eldarin and, as Flieger (2002) suggests, more separated from the light of Aman, but it is closer to the most essential, simplest roots of the Eldarin languages, and its purity in this regard allows speakers of Sindarin to channel power that Quenya's more complex morphology overshadows. It may be the case that Quenya's addition of linguistic material allows it to capture more shades of meaning, or nuance—I will not investigate such a claim here—but this ornamentation prevents it from fulfilling the role as language of power that Sindarin holds in Middle-Earth. Perhaps the implication is that magical power is not tied to Flieger's light, but is invested in the literal words of Elvish and the roots of language in Middle-Earth, and Sindarin's relative lack of obfuscation of these roots allows it to be a more efficient instrument of this power.

These brief attempts to apply my findings show just the beginning of possibilities for applications of lexicostatistics to Tolkien. In addition to these points, there are, of course, other questions to examine. For example, a future study could calculate the estimated duration of separation between Quenya and

Common Eldarin and Sindarin and Common Eldarin independently in order to determine which language changes more quickly, which could develop the idea that Quenya is a language adept at conserving information (i.e. a language of lore). However, such a study would need access to data that identifies specific Eldarin words as the most commonly used vocabulary items for a given concept; just as Sindarin has two words for 'tooth', *anc* and *carag*, each has an Eldarin root, ÁNAK and KÁRAK respectively; which of these roots should be taken to be the most common root for 'tooth' in the Swadesh list? Such questions are not easily addressed and complicate a potential study. The value in such a study would be to see *which* language changes more slowly. Here, I have used the measure of lexical distance to approximate which language is "closer" to Common Eldarin, but an independent glottochronological analysis of each language would allow for a more accurate measure of difference that would avoid the difficulties in comparing Common Eldarin roots—which were *not* the actual words used in Common Eldarin conversation, but rather are reductionist shorthand to distill more complex forms into basic elements—to Quenya and Sindarin words, which are attested. Additionally, the data from the quantitative analysis of Tolkien's languages could be used to further explore questions raised earlier in this thesis, such as the application of Austin's speech act theory in light of the idea that Sindarin is a language of magical performatives. In the case of light, language, and Tolkien, it is difficult to determine where literality ends and metaphor begins, and, just as Quenya and Sindarin are uniquely suited to perform different social and magical functions, the continued marriage of quantitative linguistic methodology with literary analysis of Tolkien allows us to approach these questions from multiple complementary angles.

Appendix

The various tables of this appendix use the International Phonetic Alphabet, not the traditional orthography for Elvish languages, aside from the English glosses. Long vowels are represented with acute accents rather than with the standard <ː>, meaning [aː] is represented as <á>. <?> or <-> represents a lack of information. The conventional < Ø> indicates linguistic zero, the lack of sound.

Table 10 – Quenya and Sindarin reflexes of Proto-Eldarin consonants based on Allan (1978a)[8]

Proto-Eldarin	Quenya	Sindarin
p	p	p~b
t	t	t~d
tp	?	f
tt	tt	?
ts	ts	?
tk	?	x
ty	ty/t	k~g
ky	ty/t	k~g
k	k	k~g
kw	kw	p~b
b	v	b~v
d	l	d~ð
dy	ly	g~Ø
gy	ly	g~Ø
g	Ø	g~Ø
gw	v/w/u	b~v
f	f	f
?	s	?
s	s~r	s~h/Ø
ss	ss	ss
sr	r̥	r̥ ~r
sl	l̥	l̥ ~l
sw	ʍ	ʍ ~w
sy	ɕ	h/Ø

[8] <~> means that the phones are in alternation with each other due to processes of lenition. </> is used when this alternation is due to other factors.

xy	ɛ	h/Ø
x	h	h/Ø/x
xt	ht	e?/i?
xw	hw	f
h	h	h/Ø
w	v/w/u	gw~w/u
r	r	r
rp	rp	rf
rt	rt	r?
rk	rk	rx
rb	rv	rv/-rf
rd	rd	rð/-r?
rg	r	r/-rx
l	l	l
lp	lp	lvb
lt	lt	lvd
lk	lk	l
lb	lv	lvf
ld	ld	ldð
lg	l	l
j	j	i/e
mb	m/mb	b
nd	n/nd	d
ndj	nj/ndj	gi~e/i
?gj	nj/ndj	gi~e/i
?g	n/ng	g
?gw	nw/ngw	b
m	m	m~v
n	n	n
nb	?	mb
nr	nr	nð
nn	nn	nn
nj	nj	ng
?	n/ng	ng
mp	mp	m/mm/-mp
nt	nt	nn/-nt
ntj	ntj	ng/-nk
nkj	ntj	ng/-nk
?k	nk	ng/-nk
?kw	nkw	m/mm/-mp

mb	mb	m/mm
nd	nd	nn/nd
ndr	ndr	ndr
ndj	ndj	ng
ngj	ndj	ng
?g	ng	ng
?gw	ngw	m/mm

Table 11 – Swadesh list for Sindarin and Quenya

#	Meaning	Sindarin	Quenya	Cognate status	Shared Eldarin root
1.	I	ni	ni	Y	-
2.	thou (informal sg. 'you')	gin	tje	N	
3.	we	men	me	Y	me
4.	this	sen	sina	Y	si
5.	that	taw	tana	Y	ta
6.	who	man	man	Y	-
7.	what	man	mana	Y	-
8.	not	ú	lá	(see footnote 6)	
9.	all	pán	illi	N	
10.	many	laew	limbe	Y	li
11.	one	min	mine	Y	mini
12.	two	tad	atta	Y	tata
13.	big	daer	hoa	N	
14.	long	and	anda	Y	ánad
15.	small	niben	níka	Y	nip
16.	woman	rín	nís	N	
17.	man	benn	nér	N	
18.	person	pen	kwén	Y	kwen
19.	fish	lim	lingwe	Y	liw
20.	bird	aew	aiwe	Y	ajw
21.	dog	hu	huo	Y	khug
22.	louse	-	-		
23.	tree	galað	alda	Y	ƞalat
24.	seed	ereð	erde	Y	eréd
25.	leaf	las	lasse	Y	las
26.	root	θond	sundo	Y	thud
27.	bark	ríf	parma	N	
28.	skin	flád	helma	N	

29.	flesh	r̥aw	r̥áve	Y	sraw
30.	blood	sereg	serke	Y	serek
31.	bone	-	akso		
32.	grease	-	-		
33.	egg	-	-		
34.	horn (animal)	rafn	rasse	Y	ram
35.	tail	-	-		
36.	feather	-	kwesse		
37.	hair	find	fine	Y	phin
38.	head	dól	cár	N	
39.	ear	l̥aw	lar	Y	slas
40.	eye	hen	hen	Y	khen
41.	nose	bund	nengwe	N	
42.	mouth	-	anto		
43.	tooth	ank	nelet	N	
44.	tongue	lam	lamba	Y	lab
45.	claw	gamp	nappa	N	
46.	foot	tál	tál	Y	tal
47.	knee	-	occa		
48.	hand	cam	ma	N	
49.	belly	-	-		
50.	neck	iaeθ	jat	Y	jak
51.	breasts	-	-		
52.	heart	gúr	hón	N	
53.	liver	-	-		
54.	drink	soga	suk	Y	suk
55.	eat	mad	mat	Y	mat
56.	bite	nag	nak	Y	nak
57.	see	keni	ken	Y	ken
58.	hear	laθ	l̥ar	Y	las
59.	know	ista	ista	Y	ith
60.	sleep	-	-		
61.	die	gwanna	fir	N	
62.	kill	dag	nahta	N	
63.	swim	-	-		
64.	fly	renia	wil	N	
65.	walk	pada	vanta	N	
66.	come	tol	tul	Y	tul
67.	lie	-	kaita		
68.	sit	hav	har	Y	kham
69.	stand	-	tar		

70.	give	anna	anta	Y	an
71.	say	ped	kwet	Y	kwet
72.	sun	anor	anar	Y	anár
73.	moon	iθil	isil	Y	ithil
74.	star	él	elen	Y	elen
75.	water	lorn	nen	N	
76.	rain	ross	miste	N	
77.	stone	hond	ondo	Y	gond
78.	sand	liθ	litse	Y	lit
79.	earth	amar	kemen	N	
80.	cloud	fain	fanja	Y	span
81.	smoke	osp	-		
82.	fire	naur	ur	Y	nar
83.	ash	liθ	-		
84.	burn	-	usta		
85.	path	imrad	tie	N	
86.	mountain	aegas	oron	N	
87.	red	born	karne	N	
88.	green	kalen	laika	N	
89.	yellow	malen	malina	Y	smal
90.	white	brassen	ninkwe	N	
91.	black	morn	more	Y	mor
92.	night	daw	lóme	N	
93.	hot	born	saiwa	N	
94.	cold	helx	ringa	N	
95.	full	pant	kwanta	Y	kwat
96.	new	kir	vinja	N	
97.	good	maer	mára	Y	mag
98.	round	korn	korna	Y	kor
99.	dry	afarx	parka	Y	anpárak
100.	name	eneθ	esse	Y	éneth

Table 12 – Most to least similar common words in Sindarin and Quenya

Meaning	Sindarin	Quenya	Lexical distance
I	ni	ni	0
who	man	man	0
eye	hen	hen	0

foot	tál	tál	0
know	ista	ista	0
round	korn	korna	0.2
what	man	mana	0.25
one	min	mine	0.25
long	and	anda	0.25
hair	find	fine	0.25
see	keni	ken	0.25
give	anna	anta	0.25
sun	anor	anar	0.25
moon	iθil	isil	0.25
black	morn	more	0.25
we	men	me	0.333333333
dog	hu	huo	0.333333333
eat	mad	mat	0.333333333
bite	nag	nak	0.333333333
come	tol	tul	0.333333333
sit	hav	har	0.333333333
yellow	malen	malina	0.333333333
leaf	las	lasse	0.4
blood	sereg	serke	0.4
this	sen	sina	0.5
that	taw	tana	0.5
bird	aew	aiwe	0.5
seed	ereð	erde	0.5

tongue	lam	lamba	0.5
stone	hond	ondo	0.5
fire	naur	ur	0.5
full	pant	kwanta	0.5
tree	galað	alda	0.6
root	θond	sundo	0.6
horn (animal)	rafn	rasse	0.6
sand	liθ	litse	0.6
cloud	fain	fanja	0.6 [9]
fish	lim	lingwe	0.666666667
ear	l̥aw	lar	0.666666667
hear	laθ	l̥ar	0.666666667
two	tad	atta	0.75
person	pen	kwén	0.75
flesh	r̥aw	r̥áve	0.75
neck	iaeθ	jat	0.75
drink	soga	suk	0.75
say	ped	kwet	0.75
star	él	elen	0.75
good	maer	mára	0.75
name	eneθ	esse	0.75
many	laew	limbe	0.8
small	niben	níka	0.8

[9] This figure treats the nasal and palatal as two distinct phonemes. If it were to be treated as [ɲ], the palatal nasal, the lexical distance would be 0.5.

dry	afarx	parka	0.8

Table 13 – Sindarin and Quenya lexical distance compared to Eldarin roots

Eldarin root	Meaning	Sindarin	Distance	Quenya	Distance	More conservative reflex
ajw	bird	aew	0.333	aiwe	0.500	Sindarin
an	give	anna	0.500	anta	0.500	Equal
ánad	long	and	0.500	anda	0.750	Sindarin
anár	sun	anor	0.250	anar	0.250	Equal
anpárak	dry	afarx	0.714	parka	0.714	Equal
elen	star	él	0.750	elen	0.000	Quenya
éneth	name	eneθ	0.600	esse	1.000	Sindarin
eréd	seed	ereð	0.500	erde	0.500	Equal
gond	stone	hond	0.250	ondo	0.500	Sindarin
ith	know	ista	0.500	ista	0.500	Equal
ithil	moon	iθil	0.400	isil	0.400	Equal
jak	neck	iaeθ	0.750	jat	0.333	Quenya
ken	see	keni	0.250	ken	0.000	Quenya
kham	sit	hav	0.500	har	0.500	Equal
khen	eye	hen	0.250	hen	0.250	Equal
khug	dog	hu	0.500	huo	0.500	Equal
kor	round	korn	0.250	korna	0.400	Sindarin
kwat	full	pant	0.750	kwanta	0.333	Quenya
kwen	person	pen	0.500	kwén	0.250	Quenya
kwet	say	ped	0.750	kwet	0.000	Quenya
lab	tongue	lam	0.333	lamba	0.500	Sindarin

las	leaf	las	0.000	lasse	0.400	Sindarin
las	hear	laθ	0.333	l̥ar	0.667	Sindarin
li	many	laew	0.750	limbe	0.600	Quenya
lit	sand	liθ	0.333	litse	0.400	Sindarin
liw	fish	lim	0.333	lingwe	0.500	Sindarin
mag	good	maer	0.500	mára	0.750	Sindarin
mat	eat	mad	0.333	mat	0.000	Quenya
me	we	men	0.333	me	0.000	Quenya
mini	one	min	0.250	mine	0.250	Equal
mor	black	morn	0.250	more	0.250	Equal
nak	bite	nag	0.333	nak	0.000	Quenya
ŋalat	tree	galað	0.400	alda	0.600	Sindarin
nar	fire	naur	0.250	ur	0.667	Sindarin
nip	small	niben	0.600	níka	0.750	Sindarin
phin	hair	find	0.750	fine	0.750	Equal
ram	horn (animal)	rafn	0.500	rasse	0.600	Sindarin
serek	blood	sereg	0.200	serke	0.400	Sindarin
si	this	sen	0.667	sina	0.500	Quenya
slas	ear	l̥aw	0.750	lar	0.500	Quenya
smal	yellow	malen	0.600	malina	0.667	Sindarin
span	cloud	fain	0.750	fanja[10]	0.800	Sindarin
sraw	flesh	r̥aw	0.500	r̥áve	1.000	Sindarin
suk	drink	soga	0.750	suk	0.000	Quenya
ta	that	taw	0.333	tana	0.500	Sindarin

[10] As discussed previously, the nasal in *fanja* could be treated as [nj] or [ŋ]. If treated as [ŋ], the distance would be 1.000.

tal	foot	tál	0.333	tál	0.333	Equal
tata	two	tad	0.500	atta	0.500	Equal

Bibliography

Allan, J. (1978a). Proto-Eldarin consonants. In J. Allen (Ed.), *An introduction to Elvish* (pp. 128-142). Glenfinnan, UK: Bran's Head.

Allan, J. (1978b). Quenya grammar & dictionary. In J. Allan (Ed.), *An introduction to Elvish* (pp. 3-44). Glenfinnan, UK: Bran's Head.

Allan, J. (1978c). Sindarin grammar & dictionary. In J. Allan (Ed.), *An introduction to Elvish* (pp. 47-71). Glenfinnan, UK: Bran's Head.

Fauskanger, H. (2008). English – Quenya [dictionary]. Retrieved from http://www.ambar-eldaron.com.

Flieger, V. (2002). *Splintered light: Logos and language in Tolkien's world*. Kent, OH: Kent State University.

Fortson, B. W., IV. (2010). *Indo-european language and culture: An introduction* (2nd ed.). Chichester, UK: Wiley-Blackwell.

Gilson, G., & Welden, B. (1978). Proto-Eldarin vowels: A comparative survey. In J. Allan (Ed.), *An Introduction to Elvish* (pp. 107-127). Glenfinnan, UK: Bran's Head.

Haugen, E. (1990). Danish, Norwegian and Swedish. In B. Comrie (Ed.), *The world's major languages* (pp. 157-179). Oxford, UK: Oxford University Press.

Hostetter, C. (2007). Tolkienian linguistics: The first fifty years. *Tolkien Studies, 4*, 1-46. doi: 10.1353/tks.2007.0022

Knut, B., & Vogt, H. (1962). On the validity of glottochronology. *Current Anthropology, 3*(2), 115-153. Retrieved from http://www.jstor.org

Lees, R. (1953). The basis of glottochronology. *Language, 29*(2), 113-127.

Millar, R. M. (2007). *Trask's historical linguistics* (2nd ed.). London, UK: Hodder Arnold.

Minkova, D. (2004) Philology, linguistics, and the history of [hw]~[w]. In A. Curzan & K. Emmons (Eds.), *Studies in the history of the English language II: Unfolding conversations.* (pp. 7-46). Berlin: de Gruyter.

Noel, R. (1974). *The languages of Tolkien's Middle-earth*. Boston, MA: Houghton Mifflin.

Petroni, F., & Serva, M. (2008). Language distance and tree reconstruction. *Journal of Statistical Mechanics: Theory and Experiment, 8.* doi:10.1088/1742-5468/2008/08/P08012

Petroni, F., & Serva, M. (2011). Automated word stability and language phylogeny. *Journal of Quantitative Linguistics, 18*(1), 53-62. doi:10.1080/09296174.2011.533589

Salo, D. (2004). *A gateway to Sindarin.* Salt Lake City, UT: University of Utah Press.

Sjoberg, A., Sjoberg, G., Swadesh, M., & Sreekantaiya, T.N. (1956). Problems in glottochronology. *American Anthropologist, 58*(2), 296-308. doi: 10.1525/aa.1956.58.2.02a00070

Solopova, E. (2009). *Language, myths, and history: An introduction to the linguistic and literary background of J.R.R. Tolkien's fiction*. New York, NY: North Landing Books.

Starostin, G. (2010). Preliminary lexicostatistics as a basis for language classification: A new approach. *Journal of Language Relationship, 3*, 79-116. Retrieved from http://www.jolr.ru

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society, 96*(4), 452-463. Retrieved from http://www.jstor.org

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics, 21*(2), 121-137. Retrieved from http://www.jstor.org

Tolkien, J.R.R. (1987). *The lost road and other writings.* C. Tolkien (Ed.). New York, NY: Random House.

Tolkien, J.R.R. (1993). *The war of the jewels.* C. Tolkien (Ed.). Boston, MA: Houghton Mifflin.

Tolkien, J.R.R. (2001). *The Silmarillion* (2nd ed.). C. Tolkien (Ed.). New York, NY: Random House.

Tolkien, J.R.R. (2000). The letters of J.R.R. Tolkien. H. Carpenter & C. Tolkien (Eds.). New York, NY: Houghton Mifflin.

Tolkien, J.R.R. (2005). *The lord of the rings* (5th ed.). Boston, MA: Houghton Mifflin.

Tolkien, J.R.R. (2010). Quenya phonology. C. Gilson (Ed.). *Parma Eldalamberon,* 19, 1-108.

Wiese, R. (1996). *The phonology of German.* Oxford, UK: Oxford University Press.

Academic Vita

Christian D. Brendel

cdbrendel@gmail.com

Education

Pennsylvania State University, Berks Campus, Reading, Pennsylvania
Major: Communication Arts and Sciences
Honors in Communication Arts and Sciences
(August 2010 – May 2014)

University of Leeds, Leeds, England
Linguistics and Phonetics, School of Modern Languages and Cultures
(January 2013 – June 2013)

Honors & Awards

- The Beinecke Scholarship
- Young Investigator Award
- The Evan Pugh Scholar Award (3rd and 4th year)
- The University of Leeds Travel Award
- Ambassador Travel Grants (Taiwan, England, Ireland, & South Africa)
- The Boscov Scholarship
- Commonwealth Scholarships (Italy, Taiwan, & Ireland)
- First-Year Advocate Award
- The President Sparks Award
- Jackie Ciferri Award for First-Year Writing
- The President's Freshman Award

Memberships

- Schreyer Honors College
- Lambda Pi Eta honor society
- Phi Kappa Phi honor society

Competitive Internships

- Intern, Natural History Research Experiences, Smithsonian National Museum of Natural History, Washington, DC
  (June 2013 – August 2013)

- Intern, Assessment Development, Educational Testing Service, Princeton, New Jersey
  (July 2012 – August 2012)

Academic Services & Support

- Writing Tutor, Penn State Berks Writing Center, Reading, Pennsylvania
  (January 2012 – May 2014)

- Lion Ambassador Tour Guide, Penn State Berks, Reading, Pennsylvania
  (May 2011 – May 2014)
  - Vice President of Special Events (August 2011 – May 2012; August 2013 – May 2014)

- President, Honors Club, Penn State Berks, Reading, Pennsylvania
  (August 2012 – May 2014)

- Peer Mentor, Penn State Berks, Reading, Pennsylvania
  (August 2011 – December 2013)

- New Student Orientation Leader, Penn State Berks, Reading, Pennsylvania
  (August 2011 – December 2013)

- Peer Reviewer for *Young Scholars in Writing*
  (March 2012 – July 2013)

- Peer Reviewer for *The Undergraduate Journal of Service Learning and Community-Based Research*
  (March 2012 – July 2012)

International Experiences

- Semester abroad at the University of Leeds, England
  (January 2013 – June 2013)

- English 297H "Writing and Communicating Health in South Africa", involving one-week study in Johannesburg, South Africa
  (March 2014)

- English 300M "Passages", involving one-week study in Donegal and Dublin, Ireland
  (November 2013)

- English 297H "Constructing Taiwan's Nature", involving one-week study and teaching English to elementary school children in Yilan, Taiwan (March 2012)

- English 297H "Monsters of Venice"
  (December 2010 – January 2011)

Publications and Presentations

- "A Kaleidoscope of Perspectives" panel, National Conference on Peer Tutoring in Writing (Tampa, FL), 2013

- "Flowing Dunes: Using Blogs Across Institutions for Tutor Development" panel, International Writing Center Conference (San Diego, CA), 2012

- "Tutoring between Language with Comparative Multilingual Tutoring", *The Writing Center Journal*, 32.1, pp. 78-91, 2012.

- "Why Not Wiglaf?", Higher Education Council of Berks County Conference, Alvernia University, 2011.

- "The Language Demons in *Paradise Lost",* Undergraduate Conference in Medieval and Early Modern Studies, Moravian College, 2010

Civic Engagement

- Organizer of/participant in community service activities, Honors Club, Penn State Berks, Reading, Pennsylvania
  (August 2010 – May 2014)

- Community Service officer, Honors Club, Penn State Berks, Reading, Pennsylvania
  (August 2010 – May 2011)

- Community Service Pre-Orientation Summer program, Penn State Berks, Reading, Pennsylvania
  (July 2010)