A Macro-Etymological Analysis of James Joyce's *A Portrait of the Artist as a Young Man*

Jonathan Reeve

1 May 2015

The English language is a palimpsest, bearing traces of the languages it has contacted. French, Latin, Ancient Greek, and Irish are among the languages that have contributed words to English, and these ancestor languages comprise modes of expression whose every word recalls the contexts of their acquisition. When a writer chooses the word "chew" over "masticate," or "enchantment" over "spell," what does that decision indicate? How can we measure these stylistic vectors? This study uses a computational analysis of the etymologies of words in James Joyce's novel *A Portrait of the Artist as a Young Man* in order to identify etymological registers, generic resonances, and levels of formality. In particular, this study will attempt to measure the maturing language of this novel's protagonist Stephen Dedalus through his use of Latinate words, and to identify ways in which this novel's macro-etymological signals reflect its structural elements.

The works of James Joyce are ideal for macro-etymological analysis. Joyce was famously multilingual, and many see his novels as a crescendo of linguistic experiments. In the words of Laurent Milesi, "Joyce's *oeuvre* is best seen as constantly trying to inform an evolutive linguistic poetics" (1). *Finnegan's Wake*, the culmination of his career in literary experimentation, is arguably unparalleled in its paranomasia and polysemy; in its composition, Joyce employed word roots from forty languages[1]. But this impulse was present in Joyce's early works, as well: his words are deliberately chosen to suggest their ancestors and cognates. They are serio-comic puns made to extend along etymological axes to new meanings in other languages.

Joyce himself was keenly interested in etymology. In his early critical essay "The Study of Languages," he argues that "in the history of words there is much that indicates the history of men, and in comparing the speech of to-day with that of years ago, we have a useful illustration of external influences on the very words of a race" ("The Study of Languages" 15). Joyce's interest in etymology was that of the application of word history to English usage. He argues for

[1] These languages are listed at the end of Joyce's manuscript for the *Wake* (Wales 31).

the study of Latin, of which "a careful and well-directed study must be very advantageous," because it "acquaints us with a language, which has a strong element in English, and thus makes us know the derivations of many words, which we then apply more correctly and which have therefore a truer meaning for us" (16). This itself may be an etymological pun, since the word *etymology* is derived from the Greek *etumon* for "true." It follows that, by studying the etymologies of Joyce's words, we might discover more of the diversity of what Joyce considered to be "truer."

In *Stephen Hero*, an early version of *Portrait*, Stephen Dedalus is described as having "read Skeat's Etymological Dictionary by the hour" (*Stephen Hero* 32). Although purely autobiographical readings of the two novels are problematic, we may safely assume that Joyce had also read and loved this work. In an etymological reading of the *Dubliners* story "Ivy Day in the Committee Room," Michael Brian argues that Joyce had a such a "detailed and profound knowledge" of Skeat's dictionary, and that it had such an influence on this story that "one could say [it] is written in Skeatish" (220). Stephen Whittaker takes it as obvious that Joyce was intimately familiar with Skeat, to the extent that he is more interested in the question of whether Joyce worked from the third or fourth edition of the dictionary (178).

In *Portrait*, Stephen routinely muses about words, considering their sounds, shapes, and beauty. "Suck," Stephen considers "a queer word" (8), but "wine" he thinks "a beautiful word" (39). Seeing the word "fœtus" carved into a desk "startle[s] his blood," (75) but upon hearing Cranly say "mulier cantat," he remarks on the "soft beauty of the Latin word" (205). It is this logophilia that justifies, in part, the following quantitative methodology, even at the risk of decontextualizing individual words. "One difficulty in esthetic discussion," Stephen seemingly cautions us, "is to know whether words are being used according to the literary tradition or according to the tradition of the marketplace" (157). This is one of the difficulties of computational literary criticism, as well: the so-called "bag of words" model of digital text analysis cannot sufficiently account for context. Conversely, Joyce's attention to words and their histories valorizes an investigation such as this.

Marjorie Howes argues that Joyce "consistently embedded the complexities of colonialism and nationalism in particular words," and cites his use of *ivory*, as a spiritual metaphor (Mary is a "tower of ivory" (Joyce, *P* 150)), a sexual image (Eileen's hands were like ivory (Joyce, *P* 37)), and a colonial commodity (255). Stephen daydreams about this word, and imagines it prisming: "The word now shone in his brain, clearer and brighter than any ivory sawn from the mottled tusks of elephants. *Ivory, ivoire, avorio, ebur*" (*P* 150). The splitting of the word into its four cognates approximately traces its etymology, from English to (Norman) French to Latin. In fact, each of these four forms for "ivory" are given both in Skeat's dictionary and in the OED in precisely this order, although this is not their direct lineage. This word history, therefore, traces a path that locates language among nations, and finds "the history of men" in "the history

of words" ("The Study of Languages" 15). This vector is also Joyce's own biographical path of exile, from Ireland to France and finally to Rome, where he first began to rework *Stephen Hero* into *Portrait*.

The narrative style of *Portrait* is another of its properties that makes it appropriate for macroanalysis. Whether called *Erlebte Rede*, or, in Flaubert's term, *le style indirect libre*, it is style in which the boundaries between the narrator's language and the characters are blurred. When Wyndham Lewis disparaged Joyce's phrase "Uncle Charles repaired to the outhouse," complaining that "people *repair* to places in works of fiction of the humblest order," Hugh Kenner responded by explaining that " 'repaired' wears invisible quotation marks. It would be Uncle Charles's own word should he chance to say what he was doing" (17). Kenner thus dubbed this Joycean narrative technique the "The Uncle Charles Principle," which he defines by explaining that "[Joyce's] words are in such delicate equilibrium, like the components of a sensitive piece of apparatus, that they detect the gravitational field of the nearest person" (16). For Kenner, this style is primarily observed on the level of the individual word. "[Joyce] is not," he writes, "like Beckett, an Eiffel nor a Calder of the sentence. The single word—'repaired'; 'salubrius'—is his normal means to his characteristic effects" (20). This might be because, as Joyce was aware, the histories of each word made them richly polysemous. This property of *Portrait* is one that makes macroanalysis meaningful: the histories of the individual words aren't simply functional aspects of the language, but crucial stylistic and ontological units saturated with traces of their origins.

Since Uncle Charles himself makes only a momentary appearance in *Portrait*, a more significant effect of the Uncle Charles Principle may be observed in the language of Stephen Dedalus, whether expressed directly or through the narrator. Stephen's language, and therefore largely the language of the novel as a whole, begins with juvenile songs and ends with mature prose. The following experiment is designed to quantify that development, by analyzing each of the chapters of the novel individually. The initial hypothesis is that the macro-etymological analyzer will show an increase in proportions of words of Latinate origin throughout the course of the novel. This hypothesis is confirmed, but not without surprises.

The Experiment

The Macro-Etymological Analyzer is a web app written using a LAMP stack—Linux, Apache, MySQL, and PHP. It ingests a text, tokenizes it, and looks up each word in the Etymological Wordnet, a relational database created from Wiktionary data by the computer scientist Gerard de Melo. The program finds the first language ancestor of each word, and categorizes it according to language family. Since words of French origin and words of Latin origin often share roots—many English words come from Latin through French or Anglo-Norman—these

are grouped together into the category "Latinate," along with words of Italian or Spanish origin. Words descended from Old or Middle English, German, or Dutch are categorized as "Germanic"; words of ancient and modern Greek origin are denoted "Hellenic"; and words of Irish or Scottish origin are "Celtic." The program then determines the proportions of words of each category[2]. *A Portrait of the Artist as a Young Man* contains 90% words of Germanic origin, 5% words of Latinate origin, and less than 0.1% each of Hellenic, Slavic, Iranian, Afroasiatic, and Celtic. A further 4% of the words in the text were not found in the dictionary, many of them proper names. These data alone are not very interesting, however, since we have no control group with which to compare them. We must therefore begin by calibrating the program.

Calibration

To find significance in these etymological signals, the Macro-Etymological Analyzer was trained on genres extracted from the Brown University Standard Corpus of Present-Day American English, a much-studied linguistic corpus of approximately one million words, created in the 1950s[3]. The corpus is broken into genre categories such as "science fiction," "belles lettres," "humor," and "news." Each of these categories was extracted using the Python NLTK and analyzed. Figure 1 shows the occurrence of Latinate words in categories of the Brown Corpus. The genres are divided fairly cleanly between fiction and non-fiction, with the fiction genres "adventure" and "romance" on the low end of the spectrum, and the non-fiction genres "learned" and "government" on the high end. Strikingly, the genres "Lore" and "Religion," which are arguably of ambiguous fictionality, fall in the middle. "Science Fiction," which is probably the most non-fictional of the fiction genres, lies in the same quadrant, and exhibits the highest proportion of Latinate words of a fictional genre. Based on this calibration, we might say that high proportions of Latinate words (hereafter "L scores") in *Portrait* would have a good chance of exhibiting styles similar to learned text, official documents, or non-fiction.

Among proportions of Hellenic words, as shown in Figure 2, the picture is similar, but with a few key differences. Here, "religion" has a higher rank, and "government" a lower. Since Hellenic words represent such a tiny percentage of any given text, however—a total of 66 words for *Portrait*—we cannot treat measurements of this category as equally statistically significant. The same is even more true for proportions of words of Celtic origin, since only a single word was detected in that category. Germanic etymologies were inversely correlated

[2]At the moment, these are proportions of the total tokens, but a future version of this program will calculate proportions of the types.

[3]An important early study of this corpus Kučera and Francis, *Computational Analysis of Present-day American English* (**???**); a more recent study is Geoffrey Leech, *Change in Contemporary English: A Grammatical Study* (**???**).

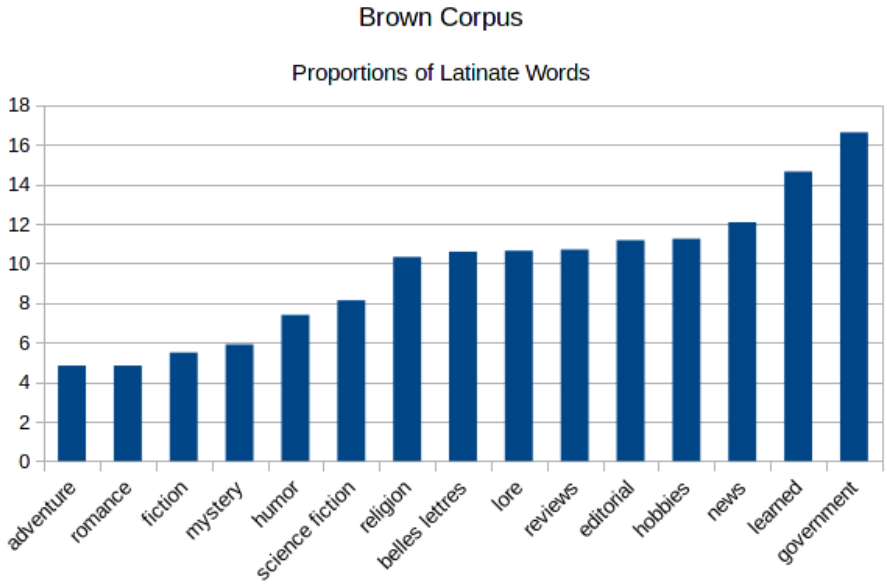

Figure 1: Brown Corpus: Latinate Words

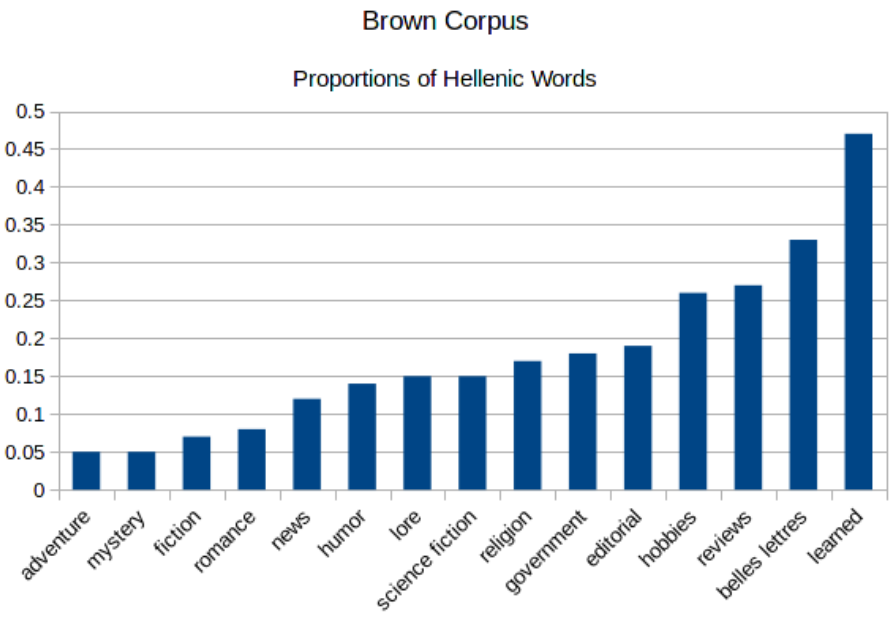

Figure 2: Brown Corpus: Hellenic Words

with Latinate etymologies, so these values are already roughly represented by L scores. Each of these categories deserves an in-depth discussion.

Languages

Latinate

The calibration experiments performed above suggest that high proportions of Latinate words are correlated with non-fiction and formal or authoritarian language. In part, this can be explained by the history of the introductions of Latinate words to English. Directly following the Norman Conquest of 1066, French became the language of aristocracy, and where French words entered English, it was often in this domain. A classic example is that names of animals—*cow*, *pig*, and *deer*, for instance, are almost all of Old English inheritance, while the names of those meats at the table–*beef*, *pork*, and *venison*—are of French (**???** 255). The English-speaking lower classes would be more likely to be in contact with the animals themselves, while the French-speaking upper classes would be likelier to be concerned with the commodity[4].

As previously discussed, the hypothesis for the analysis of *Portrait* was that there would be an increase in the L scores across chapters in the novel. Figure 3 shows that this hypothesis is partially confirmed. There is a rise in the proportions of Latinate words over chapters 1, 2, and 3, which would seem to correlate with the maturation of Stephen's thought and speech. The L score plateaus or drops in chapters 4 and 5, however. How might this be interpreted?

To answer this question, it is necessary to conduct a more granular analysis. Figure 4 shows the text is divided into sections—segments of chapters Joyce separates with rows of asterixes[5]. The L scores for these divisions exhibit much less of a simple progression from low to high. Where the climax of the chapter-based analysis seemed to be in Chapter 3, the climax here appears to be Chapter 4, Part 1. With the exception of Chapter 2, the longest and only five-section chapter, the highest L scores for each chapter come in the first section. The final sections of each chapter are among the lowest in L scores.

Seen broadly, there is a pattern here suggestive of a what Riquelme calls a "structural rhythm"—a repeating sawtooth shape. A number of critics have noticed this cyclical structure. Sidney Bolt describes it thus:

> At the beginning of each chapter Stephen is presented as the subject of a distressing tension, which develops to a crisis leading to a resolution. At the beginning of the next chapter, however, this resolution

[4]John Paul Riquelme names and discusses each of these parts in "The Parts and the Structural Rhythm of *A Portrait*, ("Structural Rhythm" 307).

[5]John Paul Riquelme names and discusses each of these parts in "The Parts and the Structural Rhythm of *A Portrait*, ("Structural Rhythm" 307).

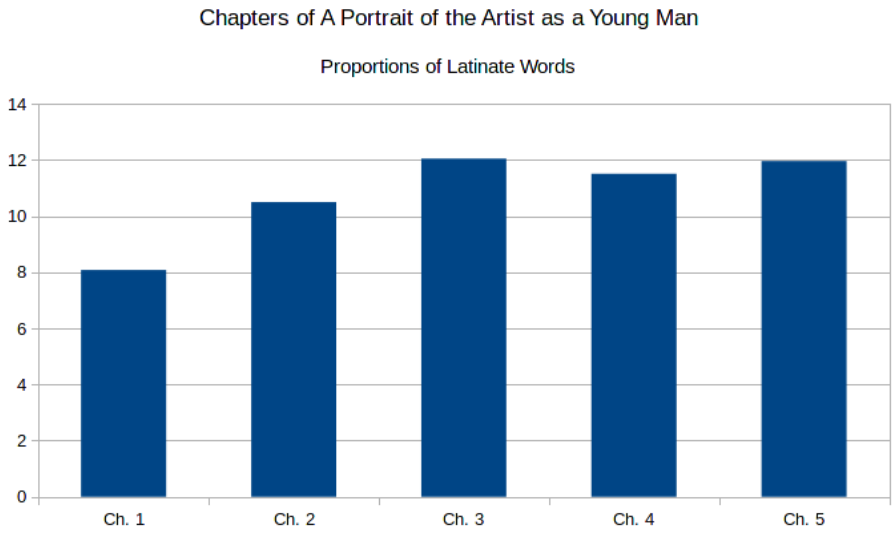

Figure 3: Chapters, L Scores

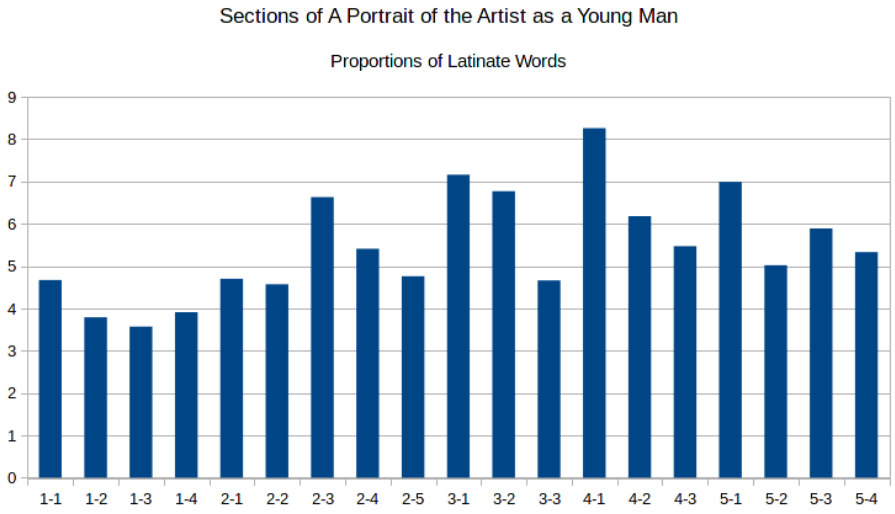

Figure 4: Sections, L Scores

> is seen to have produced a new tension, and the process is continued in a new form. This wave-like, pulsating movement is characteristic of every scene. (63)

Thomas Connolly calls this form a play between spiritual and corporeal forces. "Each [of these forces] nullifies the other," he argues, "and a nexus results until the aesthetic perception of the beautiful breaks the knot and kinesis yields to stasis" (22). Diane Fortuna describes these cycles in terms of labyrinth imagery and the Dedalus myth, and adds that "aside from the initial subsection of *Portrait*, each of the subsequent 18 divisions of the novel presents at least one image of rolling, cyclical, or circling motion" (197). Fortuna's observation could be read as an approximate description of the rolling, cyclical etymological trends shown in Figure 4.

One seminal description of this phenomenon is David Hayman's reading of this structural oscillation as one between epiphanies and anti-epiphanies. The epiphanic moment is "a lyrical and wish-fulfilling moment during which the illusory is made to appear as immediate and valid"; it is "both art and event." These moments then engender an "anti-aesthetic impulse to action" (164–5). While the epiphany is a "vision" or "illusion," it is followed by an anti-epiphany that "show[s] Stephen to be increasingly involved with the world" (174). Riquelme calls this oscillation "a stylistic double helix," and adds that "Joyce employs the two epiphanic modes of stark realism—'the vulgarity of speech or of gesture'—and visionary fantasy … as delimiting extremes in his character" ("Styles of Realism" 119, 104). These properties—lofty visions and earthly pragmatics—map roughly to properties of L and G registers.

A closer reading of these sections might be more useful than a simple mapping of criticism to macro-etymological data, however. For that, we must make the analysis even more granular. The section with the highest L score is 4.1, which Riquelme titles "Spiritual Discipline." Grant Redford, for one, claims that this is the climactic section of the novel (108). Within this section, the highest L score can be found in the second quarter. Here is a single sentence excerpted from that subsection:

> The imagery through which the nature and kinship of the Three Persons of the Trinity were darkly shadowed forth in the books of devotion which he read—the Father contemplating from all eternity as in a mirror His Divine Perfections and thereby begetting eternally the Eternal Son and the Holy Spirit proceeding out of Father and Son from all eternity—were easier of acceptance by his mind by reason of their august incomprehensibility than was the simple fact that God had loved his soul from all eternity, for ages before he had been born into the world, for ages before the world itself had existed. (124-5)

This passage is verbose, florid, and multi-syllabic; its subject matter is religious, authoritarian, and deathly serious. Compare that with the passage with the lowest L score, at the end of section 1.3, "Christmas Dinner," when the argument about Parnell becomes heated. Mr. Casey's livid yet comic remark here neatly illustrates the Germanic register used in this section: "She stuck her ugly old face up at me when she said it and I had my mouth full of tobacco juice. I bent down to her and *Phth!* says I to her like that" (30). With the notable exception of "tobacco," which is ultimately descended from an indigenous Haitian language, most of these words are monosyllabic and of Germanic origin. The rhythm here is faster, and the tone lighter. There is a certain playfulness evident in the onomatopoeia *Phth!*, a kind of neologism which we shall see is characteristic of Germanic Joyceanisms.

Germanic

While the proportions of words of Germanic origin in *Portrait* are, roughly speaking, inversely proportional to those of Latinate origin, they warrant discussion. Words of Germanic origin are frequently monosyllabic, polosive, and evoke raw, unfiltered speech that is often undecorated with euphemism and social formality. When Cantwell says "He'd give you a toe in the rump for yourself," Stephen thinks, "that was not a nice expression" (7). What if Cantwell had used the French-derived synonym "derrière," or the Latin-derived "posterior"? That might not still be a nice place to have a toe, but the expression would be more polite. "Rump" is "not a nice expression" because it bears the resonances of the Germanic register.

Early in the novel, young Stephen overhears someone use the word "suck," and thinks, "suck was a queer word … the sound was ugly" (8). This passage, and indeed, this "queer word" has been much-discussed, most notably in Derek Attridge's study (59). This may also have been what H.G. Wells had in mind when he accused Joyce of having a "cloacal obsession":

> He would bring back into the general picture of life aspects which modern drainage and modern decorum have taken out of ordinary discourse and conversation. Coarse, unfamiliar words are scattered about the book unpleasantly… (Wells, quoted in Deming 86)

If we remember that Stephen compares the sound of "suck" to that of "dirty water" going down the drain, modern drainage is literally that which creates this "coarse, unfamiliar" word sound. Wells's critique highlights the reason why passages like this one were so "coarse" for readers contemporary with Joyce—the sounds and registers of their words. In fact, *Portrait* was rejected by early English publishers on this basis. In a reader's report for the publishers Duckworth & Company, Edward Garnett calls the novel "too discursive, formless, unrestrained," because "ugly things, ugly words, are too prominent" (Deming

81). Are words like "suck" ugly because they belong to the Germanic register, and carry those associations?

Stephen's thoughts about the word "suck" begin a onomatopoetic theme that is chiefly associated with words of Germanic origin. Stephen later explains the word "kiss" in onomatopoetic terms—when he thinks of his mother's kiss, Stephen thinks, "her lips … made a tiny little noise: kiss" (10-11). Mr. Casey's *Phth!* falls into this same category. Jeri Johnson notices the preponderance of these words, and argues that "if Stephen could be said to have a theory of language at this point, it would be the bow-wow or onomatopoeic theory: the word for the thing imitates its actual acoustic equivalent in reality: 'suck' has its name because things that 'suck' make 'sucky sounds' " (xxvii).

In addition to their sonic associations, Germanic-derived words in Joyce have strong visual connotations. "The word was beautiful: wine," Stephen thinks. "It made you think of dark purple because the grapes were dark purple that grew in Greece" (39). Although *wine* has a distant ancestor in Latin (*vīnum*), its immediate parents are Middle and Old English. Stephen's associations are that of a certain dark purple color and Greece, which recalls the Homeric cliché that appears three times in Joyce's *Ulysses*: "*epi oinopa ponton*"—"the wine-dark sea."[6] This is important to keep in mind, since, on the same page, a Latin lesson begins, in which Father Arnall "asked Jack Lawton to decline the noun *mare*," or sea. Jack fails to decline the noun, and "could not go on with the plural," an implicit choice of the Germanic *sea* over the Latin *mare*. The first plural of *mare* is *maria*, which is also the Latin name for Mary. Does Lawton's failure to produce "Mary" in front of Father Arnall prefigure Stephen's eventual rejection of the Sodality of the Blessed Virgin, of which he was prefect? This would be a far-fetched hypothesis on the subject of any other author, but given Joyce's famous love of puzzles, it is entirely plausible, and it takes place along an etymological axis.

When Stephen is about to confess his sins in section 3.3, we see another passage with a low L score, notable for its alliteration:

> His blood began to murmur in his veins, murmuring like a sinful city summoned from its sleep to hear its doom. Little flakes of fire fell and powdery ashes fell softly, alighting on the houses of men. They stirred, waking from sleep, troubled by the heated air. (130)

The double alliterative structure here—a string of s- words interrupted by a string of f- words—recalls the verse style distinctive of Old English poems such as *Beowulf*. Furthermore, Most of these words are of Germanic origin, which lends them the immediacy that the passage requires to evoke Stephen's guilt and anxiety.

[6]This phrase is discussed at length in William Gladstone's *Studies on Homer and the Homeric Age*, where Gladstone argues on the basis of color words in Homer that the ancient Greeks lacked the ability to perceive colors like blue.

Hellenic

Portrait's words of ancient Greek origin deserve a brief discussion. Greek words are some of the more difficult to quantify, since most of the Greek loanwords in English come to us through Latin, and a few (like "alchemy") through Arabic. When classical Greek works began to be rediscovered in 1453, after Greek scholars fled Turkish-occupied Constantinople, this brought with them a number of associated loanwords. This could explain why many Greek loanwords seem at home in Aristotle or Plato—*drama, comedy,* and *pathos* recall the *Poetics*; *phenomenon, noumenon,* and *democracy* seem appropriate to a Socratic dialogue. As the analysis of the Brown Corpus hints, religious words, too, are heavily Hellenic: *angel*, *evangelist*, *hagiography*, *bible,* and so on, are all descended from Greek. We might find, therefore, that an aesthetic treatise of the kind Stephen presents in 5.1, or a religious sermon like Father Arnall's in 3.2, might contain a higher proportion of words of Hellenic origin.

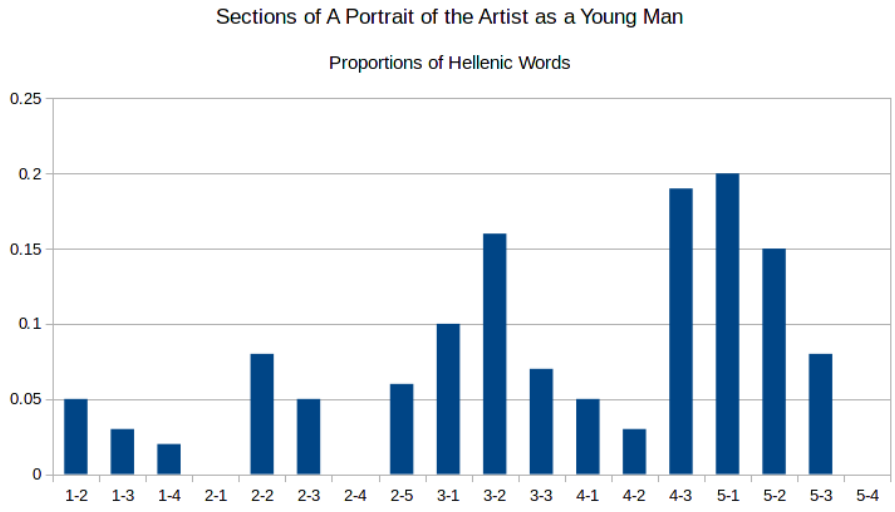

Figure 5: Sections, H Scores

Figure 5 shows that those two sections have, respectively, the first and third highest H scores of any section. Father Arnall's sermon in section 3.2 features the emotionally-charged Hellenic words *agony,* (which appears an amazing eight times in this section alone), *demon* and *zealous,* along with the more tame words *baptism, poetry*, and *eon.* In section 5.1, those words are more befitting their setting in a physics classroom—*physics*, *energy*, and *kinetic*, along with *didactic.* These categories of religious and learned language are consistent with the analysis of the Brown corpus.[7]

[7]The pattern of Hellenic words in Figure 5 also closely resembles patterns of religious images

The section with the second-highest H score is 4.3. Interestingly, this is the section where Stephen's classmates taunt him in Greek: "Stephanos Dedalos! Bous Stephanoumenos! Bous Stephaneforos!" (141). These words are a polyglot pun on his name and Greek words for a sacrificial cow adorned with a wreath (O'Hehir, *Classical* 528). However, this is not what the Macro-Etymological Analyzer is detecting—since the program doesn't recognize words of languages other than English, it treats *Bous* and *Stephanoumenos* as errors. The words of Hellenic origin in this section, then, are other English words: *ecstasy* and *antogonism,* for instance. That Joyce is using more than the usual number of Hellenic words here fits with the Dedalian myth, for on this same page, we see the epiphanic culmination of this metaphor, in the imagery of flight:

> His heart trembled; his breath came faster and a wild spirit passed over his limbs as though he were soaring sunward. His heart trembled in an ecstasy of fear and his soul was in flight. His soul was soaring in an air beyond the world and the body he knew was purified in a breath and delivered of incertitude and made radiant and commingled with the element of the spirit. An ecstasy of flight made radiant his eyes and wild his breath and tremulous and wild and radiant his windswept limbs. (141)

Moments later, there is an Icarian anti-ephiphany that risks bathos, as Stephen's thought is interrupted by the voices of a schoolmate playing in the water: "Oh, Cripes, I'm drownded!" (ibid.).

It could certainly be argued that the Hellenic words represented here are more useful to close reading than they are to distant reading. Since there are so few Hellenic words, they are statistically insignificant. However, in literary analysis, the significance of a single word could form the basis of a critical argument, while it may remain statistically uninteresting.

Celtic

Although the Macro-Etymological Analyzer identified only one word descended from the Irish language, *sugan*, the language has a deep effect on the styles of the novel. O'Hehir's *Gaelic Lexicon* identifies sixteen words of Irish descent in *Portrait* (*A Gaelic Lexicon for Finnegans Wake, and Glossary for Joyce's Other Works* 335–6). Some of O'Hehir's words, like *cool* (from *cúl*, goal) are homographs with unrelated English words, a fact that might help to explain why they cannot be found by the program[8]. Others, like "smugging," are of dubious

identified in a 1979 computational study of *Portrait* by John B. Smith **???**. In this study, Smith counts "images" that belong to certain taxonomies like "fire" and "water," and plots them according to their location in the novel. The category of "religion" aligns very roughly with the Hellenic plot in Figure 5.

[8]Word sense disambiguation is a featured planned for future versions of the Macro-Etymological Analyzer.

Irish etymology—O'Hehir supposes that this word may be derived from *smug* or *smuga*, meaning "snot, nose drip," or "slime" (ibid.). In classic Joycean fashion, this word is not glossed. It appears early in the novel, when Athy relates that the Simon Moonan and Tusker Boyle are caught "smugging" in the restroom. This is such a somber revelation that the rest of the boys are silenced by the thought, but Stephen does not understand—"what did that mean about the smugging?" he thinks (35).

Johannes Hedberg guesses that *smugging* can be traced to the Old English word *smūgan*, by way of the Middle English verb *smu en* (25), but Alarik Rynell contends that Hedberg's etymology is erroneous, and that it is more likely decended from Old English *smugge*, "a small secret place" (367). Rynell uses the English phonesthemes of *smugging* to argue that "*smugging* must indeed have seemed an appropriate colloquialism for *masturbating*." This is also Attridge's theory (63). Most others assume that it's a euphemism for homosexual play (Howes 255, for instance), although Fargnoli claims it is "entirely made up and has no established meaning" (207). The OED gives "to caress, fondle," citing another of Joyce's uses of the word in *Ulysses*, as well as the early 19th century poet Scottish poet Ebenezer Picken. That the only two citations for this sense are from Celtic writers lends some credence to the theory of Celtic etymology. On the other hand, that all of these theories assume some kind of taboo schoolyard sexual act, suggests the Germanic origin of the word, not only on the basis of G score of Brown Corpus romance texts, but also given the large number of other similar four-letter sexual words in the Germanic register.

More important to this discussion than the words themselves is the political dimension of the Irish language, especially as it existed in Ireland on the eve of independence. The revival of the Irish language was intimately associated with the nationalist movement, from which Joyce as a self-imposed exile had distanced himself both physically and intellectually, but with which he nonetheless felt some affinity. Although neither Stephen nor Joyce himself knew much Irish—Stephen stops taking Gaelic League classes after the first lesson—they seem to struggle with the dual political and linguistic dominance of Britain over Ireland. A passage that illustrates this is Stephen's conversation with the English dean over the word *tundish*. Stephen thinks:

> The language in which we are speaking is his before it is mine. How different are the words *home, Christ, ale, master*, on his lips and on mine! I cannot speak or write these words without unrest of spirit. His language, so familiar and so foreign, will always be for me an acquired speech. I have not made or accepted his words. My voice holds them at bay. My soul frets in the shadow of his language. (159)

Here, "his language" could be read as both the Dean's British English dialect and English more generally. Since Irish is the ancestral language of Ireland,

English is an "acquired speech" in this historical sense. More immediately, the tonal differences in their speech distinguish their two Englishes. Anthony Burgess has a notable phonetic interpretation of this passage, suggesting that Stephen likely chooses these four words because they are pronounced differently in British and Hibernian dialects—dipthongs instead of long open vowels, final schwas instead of retroflex Rs (28). The words are ontologically different, as well—"home" refers to different cities for the two men; "Christ" is very different for the Catholic and the Protestant; and as a student, Stephen's "mastery" is that of a subject he is taught, while the Dean's is that over people, that of a colonist and a schoolmaster.

These colonial undercurrents are useful to a discussion of etymology in Joyce, because they help to reveal choices of etymological modes as domains of nations, with histories and political uses. Joyce's decision to have Stephen's uncle Mat Davin use Irish-derived words like *camann* (from the Irish *camán*, the stick used in hurling) enforces the earlier description of him as a "young peasant" who "worshipped the sorrowful legend of Ireland" (151-2). Stephen refers to this same object with an Anglo-Saxon word when he scoffs at the most recent Irish uprising, calling it "a rebellion with hurleysticks" (169). Johnson explains that this is "a 'sneerer's' comment on the failed Fenian Rising of 1867, training having taken place not with guns but with *camann*" (274). Joyce's use of *camann* and *hurleysticks* is not interchangeable, but chosen to evoke histories, politics, and tones that each word carries.

Unknown Words

Camann was one of about four percent of the words of *Portrait* that the Macro-Etymological Analyzer failed to find. These words proved to be revealing about Joyce's style, especially concerning etymological associations. Many of the unknown words are proper names, and proper names were purposely removed from the database, as they would skew the results unnecessarily.[9] Other unknown words, however, are Joyce's inventions. Some of these are true neologisms, while others are portmanteau words or unhyphenated compound words.

While much critical attention has been paid to the neologisms in *Finnegan's Wake*, since they are undoubtedly its distinctive property, not much has been discussed regarding *Portrait*, even though at least one word, *pandybat* has its first OED citation in the novel. Regarding Joyce's work as a whole, however, Katie Wales identifies two neologistic strategies: "conversions" and "compounds" (115). "Conversion," Wales relates, "extends the semantic range of existing words by changing the grammatical function." "Compounds" refers to portmanteau words, which are littered throughout the novel. These two categories are

[9] *Portrait* would show unusually high proportions of Hellenic words, for instance, in every section where the word *Stephen* would appear. Analyses of Christian bibles showed similar results every time the word *Jesus* was mentioned, irrespective of the author's choice to use words in the Hellenic register.

roughly equivalent to Joseph Prescott's conception of Joyce's "renovation" and "innovation" (308). By "renovation," Prescott claims that Joyce "imposes on words of common currency a fresh lustre, usually the brilliance of their first years." Among the examples Prescott gives for "rennovation" is a passage from *Ulysses* where Joyce uses the word "crazy" in its etymological sense of "fractured" (309). In illustration of his category of "innovation," Prescott calls Joycean neologism "dynamic onomatopoeia," citing the "crescendo" of cat noises in *Ulysses* (311). When onomatopoeia is "triumphant" in Joyce, he argues, it constitutes the "anastomosis of style and subject." Given the etymological associations with onomatopoeia established earlier, this fusion of style and subject could be said to take place along etymological vectors, as well.

To identify a list of Joycean neologisms beyond the error words of the Macro-Etymological Analyzer, the text was run through a command-line spell-checking program, and the results sorted, with this chain of Linux commands:

```
cat portrait.txt | aspell -a | cut -d ' ' -f 2 | \
grep -v '*' | sort | uniq > misspelled.txt
```

The result was a list of words the command *aspell* determined were "misspelled." After manual curation to remove words in Latin, proper names, and real but obscure words, this became a list of Joycean terms. Most of these words are compound words, and often formed from two Germanic words, like *suddenwoven* or *rainladen*. These words exhibit several themes. First, there are color words, like *ambered*, *bloodred*, *greenwhite*, and *redeyed*, along with *hueless* and *nocolored*. Next, there are kinship terms, including *fosterbrother*, *fosterchild*, *greatgrandfather*, *halfbrother*, and *granduncle*. Another category features agrarian or pastoral terms like *cowdung*, *cowhairs*, *goatish*, *milkcar*, *boghole*, and *bogwater*. Finally, there is theme related to dirt, filth, and the street: *sootcoated*, *thumbblackened*, and *greasestrewn*. All of these categories are associated with the Germanic register.

It should perhaps not be surprising that so many of Joyce's neologisms and portmanteau words are of Germanic origin, since word compounding in this style is a feature of many modern Germanic languages, most notably modern German. In fact, many of these words, if separated into their constituent words (*great grandfather*) and translated into German, prove to be one German word (*Urgroßvater*).

Conclusions

Joyce achieves many of the narrative effects of *A Portrait of the Artist as a Young Man* through the use of etymological registers. Just as the language of his narration, according to the Uncle Charles Principle, follows the thoughts of his characters, his oscillations between Germanic and Latinate linguistic

modes mimic oscillations between epiphanic and anti-ephiphanic scenes. Macro-etymological analysis, therefore, demonstrates that it might be well-suited to become part of suite of analytic tools that can participate in the detection of structural patterns of a novel. Along with word frequency analysis, principal component analysis, metrical detection, and segmentized type/token ratio calculation, macro-etymological analysis might form a part of a greater textual analytic system that can inform and improve computational literary criticism.

Works Cited

Attridge, Derek. "'Suck Was a Queer Word': Language, Sex, and the Remainer in A Portrait of the Artist as a Young Man." *Joyce Effects*. Cambridge, UK: Cambridge University Press, 2000. 59–77. Print.

Bolt, Sidney. *A Preface to James Joyce*. Second Edi. London: Longman, 1981. Print.

Brian, Michael. "'A Very Fine Piece of Writing': An Etymological, Dantean, and Gnostic Reading of Joyce's 'Ivy Day in the Committee Room'." *ReJoycing: New Readings of Dubliners*. Lexington, KY: University Press of Kentucky, 1998. 206–227. Print.

Burgess, Anthony. *Joysprick: An Introduction to the Language of James Joyce*. New York: Harcourt Brace Jovanovich, 1973. Print.

Connolly, Thomas E. "Kinesis and Stasis: Structural Rhythm in Joyce's Portrait." *University Review* 3.10 (1966): 21–30. Print.

Deming, Robert. *James Joyce, the Critical Heritage*. New York: Barnes & Noble, 1970. Print.

Fargnoli, A Nicholas. *James Joyce A-Z*. Oxford, UK: Oxford University Press, 1995. Print.

Fortuna, Diane. "The Art of The Labyrinth." *Critical Essays on James Joyce's a Portrait of the Artist as a Young Man*. New York: G.K. Hall & Co., 1998. Print.

Hayman, David. "A Portrait of the Artist as a Young Man and L'Éducation Sentimentale: the Structural Affinities." *Orbis Litterarum* 19 (1964): 161–75. Web.

Hedberg, Johannes. "Smugging. An Investigation of a Joycean Word." *Moderna Sprak* 66 (1972): 19–25. Print.

Howes, Marjorie. "Joyce, Colonialism, and Nationalism." *The Cambridge Companion to James Joyce*. Ed. Derek Attridge. Second Edi. Cambridge, UK:

Cambridge University Press, 2006. 254–2471. Print.

Johnson, Jeri. "Introduction." *A Portrait of the Artist as a Young Man*. Oxford Wor. Oxford University Press, 2000. vii–xxxvii. Print.

Joyce, James. *A Portrait of the Artist as a Young Man*. Ed. Jeri Johnson. Oxford, UK: Oxford University Press, 2000. Print.

---. *Stephen Hero*. New York: New Directions, 1944. Print.

---. "The Study of Languages." *Occasional, Critical, and Political Writing*. Oxford, UK: Oxford University Press, 2008. Web.

Kenner, Hugh. "The Uncle Charles Principle." *Joyce's Voices*. Berkeley: University of California Press, 1978. 15–38. Web.

Milesi, Laurent. *James Joyce and the Difference of Language*. Cambridge, UK: Cambridge University Press, 2003. Print.

O'Hehir, Brendan. *A Classical Lexicon for Finnegans Wake: a Glossary of the Greek and Latin in the Major Works of Joyce incl. Finnegans Wake, the Poems, Dubliners, Stephen Hero, A Portrait of the Artist as a Young Man*. Berkeley [u.a.]: Univ. of California Press, 1977. Print.

---. *A Gaelic Lexicon for Finnegans Wake, and Glossary for Joyce's Other Works*. Berkeley: University of California Press, 1967. Print.

Prescott, Joseph. "James Joyce: A Study in Words." *Publications of the Modern Language Association of …* 54.1 (1939): 304–315. Web.

Redford, Grant H. "The Role of Structure in Joyce's Portrait." *Joyce's Portrait: Criticisms and Critiques*. Ed. Thomas E Connolly. Appleton, 1962. 102–115. Print.

Riquelme, John Paul. "Stephen Hero, Dubliners, and A Portrait of the Artist as a Young Man: Styles of Realism and Fantasy." *Cambridge Companion to James Joyce, the*. N.p., 1990. 103–130. Print.

---. "The Parts and the Structural Rhythm of A Portrait." *A Portrait of the Artist as a Young Man*. New York: W. W. Norton & Company, 2007. 307–8. Print.

Rynell, Alarik. "On the Etymology of James Joyce's Smugging." *Moderna Sprak* 66 (1972): 366–369. Print.

Smith, John B. *Imagery and the Mind of Stephen Dedalus : a Computer-Assisted Study of Joyce's A Portrait of the Artist as a Young Man*. Lewisburg, PA: Bucknell University Press, 1980. Print.

Wales, Katie. *The Language of James Joyce*. New York: St. Martin's Press, 1992. Print.

Whittaker, Stephen. "Joyce and Skeat." *James Joyce Quarterly* 24.2 (1987): 177–192. Print.