Universidad de Valladolid

FACULTAD de FILOSOFÍA Y LETRAS
DEPARTAMENTO de FILOLOGÍA INGLESA
Grado en Estudios Ingleses

TRABAJO DE FIN DE GRADO

The Grapho-Phonemics of the English Paroxytones and the fever/ever Dilemma.

Cristina Otero del Real

Tutor: Enrique Cámara Arenas
2018-2019

Abstract

The pronunciation of stressed English vowels is a complex issue for L2 speakers of the language. Vowel length contrasts are one of the core features that ensure intelligibility in the context of English as a Lingua Franca. We believe that the development of reliable grapho-phonemic rules could be useful for pronunciation teaching. Following a previous study from Cámara-Arenas, we look for the domain-specific rules that may help us account for the exceptions to the most irregular pronunciation rule. Pronunciation rule 3.2. is applicable to words like *paper* or *human,* predicting a pronunciation of the nominal version for the stressed vowel. We develop a larger set of regular words with the implementation of simple domain-specific rules applicable to each stressed vowel, bringing the regularity of this structure from 57% to 89%. Further polishing of these rules with a larger database and cleaner classification criteria is encouraged.

Keywords: grapho-phonemics, paroxytones, pronunciation rules, ELF, phonetics.

La pronunciación de la vocal tónica inglesa es una cuestión compleja para hablantes L2. El contraste de longitud en vocales es uno de los rasgos centrales que garantizan la inteligibilidad en contextos de Inglés como Lengua Franca. Creemos que el desarrollo de reglas grafo-fonémicas podría ser útil en la enseñanza de la pronunciación. Siguiendo un estudio previo de Cámara Arenas, buscamos las reglas específicas de dominio que puedan justificar las excepciones de la regla de pronunciación más irregular. La regla 3.2. se aplica a palabras como *paper* o *human,* prediciendo una pronunciación de la vocal tónica en su versión nominal. Con la implementación de simples reglas específicas de dominio, aplicables a cada vocal tónica, desarrollamos un grupo mayor de palabras regulares, desde un 57% anterior al 89%. La mejora de estas reglas con una base de datos mayor y criterios claros de clasificación sería óptima.

Palabras clave: grafo-fonémica, palabras llanas, reglas de pronunciación, ILF, fonética.

Elena, mil gracias por la paciencia,
por el 0.5% de Python que me has enseñado
y porque solo nos hemos enfadado dos veces.

TABLE OF CONTENTS

A NOTE ON FORMAT

- arrow brackets (<…>): graphemes
- hyphen: to mark an affix's position
- asterisk preceding a word: unrecorded spellings of it
- *italics* for whole word examples
- **bold** will mark the stress in a spelled word
- a period: separation of syllables, either in the word, its phonetic transcription (in either system) or its grapho-phonemic formula: *li.on*, /L AY1. AH0 N/ or /laɪ.ən/, and cv.vc
- […] examples will be numbered in-between square brackets.

We will use both ARPAbet and IPA for the phonetical transcription of words, the reason for this will be explained in the Methodology section of the article.

1. INTRODUCTION

There is not one single way in which speech should be spelled. Systems have evolved in many different ways through time for diverse reasons: either by way of a general agreement of the speaking community, a technical development, a linguistic change, a political maneuver or, for most languages, a mixture of all these. An example for English could be how the <u> in <qu> is not considered a vowel but merely a spelling convention, while in peninsular Spanish we have two different spellings for /θ/ depending on the vowel that will follow, <c> and <z>.

The English Orthography is no different in this respect than any other system, but it has been consistently considered through history to be unnecessarily chaotic and complicated. This qualification obeys to the notion that an orthographic system should only consist on a series of letter to sound combinations as unambiguous as possible. Knowing these combinations, any reader or speller might accurately interpret or represent any word regardless of its history, origin or grammatical category. This is something that English does not have. In fact, English orthography gives its users all kinds of information beyond a simple letter-to-sound correlation.

However, the issue of irregularity is not as dire as some theorists make it: English, although not a shallow orthography, has been shown to have high levels of predictability when considering grapho-phonemic information (Fry, Wijk, Treiman & Wolter, Robbins et al.). The discipline of Grapho-phonemics deals with the understanding of the language's phoneme-grapheme mappings. Which is to say how each letter can be pronounced, particularly in the sense of how a word's spelling might prompt an accurate prediction of its pronunciation.

Within the field of grapho-phonemics we can look at Venezky or Cummings, who developed rules for a better understanding of L1 spelling. But we should consider Bozman or Cámara-Arenas (2008; 2010; 2018) for an L2 perspective on grapho-phonemics being applied to teaching purposes. Pronunciation has been said to be the most basic standard for mutual intelligibility between speakers, and its teaching has historically been neglected (Nguyen; Isaacs; Jenkins 2000). The use of grapho-phonemics would provide students with a stable source of information when approaching unfamiliar words,

especially when considering that most of the input L2 students receive is in the written form (Cámara-Arenas 2008, 95).

Among the different word structures, the English paroxytone —words like *p**a**nic, f**e**ver, m**i**nion, f**o**lder* or *f**u**ture*— has been proved to be problematic in terms of the predictability of the correspondence between its spelling and its pronunciation. Not all paroxytones are problematic, however. Those paroxytones with a single consonant between their stressed vowel and the final syllable (*fever* and *future* are regular, while *panic* and *minion* are not) have received barely passing rates on their predictability This issue is not a question of full irregularity that might leave speakers no other option but rote memorization, we believe that these words are not as rebellious as it might seem from the consideration of general-systemic rules. In fact, as seen in Cámara-Arenas (2018), there is reason to believe that these irregular words might just require more specific rules.

2. STATE OF THE ART

The English writing system has been considered in need of improvement and even reformation for a long time. This impulse is not a result of recent linguistic considerations: As Cummings states, people in the Elizabethan era were already claiming that "the English spelling system was nonsensical and needed reform" (3). That is, the system needed to be more *phonetic.* We know nowadays that in fact the system serves "several goals other than that of one-to-one phoneme–letter correspondence" (Kessler and Treiman 268). Kessler and Treiman also name the main principles that have shaped the system: "conservatism, the unadapted spelling of loan words, and the representation of nonphonemic information" (269).

This "conservatism" is clear in how little the system has changed through time, and it definitely has not changed enough to quiet the voices that call for reformation; we have to thank conservatism in a way for "keeping spellings more consistent across time and across the world" (Hayes et al. 2005, 8), especially when considering the diversity of English variants around the globe.

L1 speakers, despite not having to deal with most pronunciation issues, are expected to sort out the complex spelling system of their native language before they finish middle school (Hayes et al. 2016; Pacton et al. 2018). But, for an L2 English learner, the

orthographic system is an issue that receives little attention from their instructors. In fact, pronunciation teaching is a subject that has been overlooked in most ELT or EFL contexts.

2.1. PRONUNCIATION TEACHING

Teaching innovation in the context of pronunciation has shifted between two sides along time, from the teaching of segmentals to suprasegmentals and back again. Focus in segmentals resulted in "a decrease in the significance attached to pronunciation" since such a narrow perspective "failed to have a deep impact on overall pronunciation quality" (Bakla & Demizeren 480). "CLT[1] ultimately led to a reshift in instructional emphasis from segmentals to suprasegmentals" (Isaacs 2). And while it is believed that suprasegmentals "contribute to intelligibility much more than segmentals do, some of these features are not easy to teach" (480). Other authors even question "whether they [learners] put the phonetic theory to use" (Ausín & Sutton 234). Despite the recent focus in CLT, "phonological acquisition may require recourse to a focus on forms" (DeKeyser, qtd in Isaacs 5) and "experts advocate a balanced approach to pronunciation instruction that includes the teaching of individual sounds (vowels and consonants)[2] and prosodic elements such as stress, rhythm and intonation"[3] (Burri 67). Saito establishes four dimensions in pronunciation teaching, the first being "segmental accuracy: Pronouncing new consonant and vowel sounds using L2 forms instead of using their L1 counterparts or interlanguage forms" (3). Considering pronunciation-focused corrective feedback as a "crucial component of L2 pronunciation development" (13). Other common techniques are "listening and imitating, phonetic training, minimal-pair drills, tongue twisters and reading aloud" (Nguyen 45), or repetition (Escudero-Mancebo et al.) and the "comparison between their interlanguage forms and the target ones" (Bakla & Demizeren 487).

Teaching pronunciation through grapho-phonemics is an area that has received very little attention in the literature. Pronunciation has been disregarded as an area of interest in English as a Foreign Language teaching. At the end of the twentieth century, discussion of the hypothesis of the critical period "focused on the fact that most learners past puberty

---

[1] Communicative Language Teaching

[2] These are what is normally called "segmentals".

[3] These are called "suprasegmentals".

retained a foreign accent in their L2" (Derwing 12), meaning that most studies on pronunciation teaching were dedicated "solely on accent reduction, or nativeness" (13).

However, since "the number of nonnative speakers of English is greater in number than that of native speakers" (Bakla and Demizeren 478) and with the establishment of English as an International Language, the idea of native ownership of the language has become obsolete.

As Zoghbor puts it "pronunciation has been given attention in the discussion of the global spread of English due to its strong link with accent and its potential to reflect the identities of NNSs in lingua franca settings" (832). Thus, the focus has shifted from accuracy and native-likeness to intelligibility.

2.2. English as a Lingua Franca

The relatively new idea of the Lingua Franca Core developed by Jenkins (2000) has opened a new path for investigation in pronunciation teaching in the context of ELF. The core consists on "certain phonetic features [that] are important to the maintenance of mutual intelligibility regardless of the speaker's or listener's background" (O'Neal 134) while still allowing L2 speakers to maintain their identity and accent. The LFC is also based on the speakers' ability to accommodate to their interlocutor's needs in the form of "segmental repairs and adjustments" (O'Neal 121).

Jenkins has, since then, continued on her EIL approach to pronunciation, revising the LFC features that are necessary for international intelligibility and considering its impact in terms of international testing standards (Jenkins and Leung).

Furthermore, the issue has had a strong impact not only in the assessment of L2 English, but also in perceived accent bias. Levis and Moyer point out that "NS listeners and NNS learners both assume that communication success is dependent on the NNS, and that accented speech can and should be remedied because it negatively affects communication" (285).

L2 should not need to copy the Received Pronunciation and General American native varieties which do not represent a significant amount of the global English speakers. Nowadays, it is necessary to consider a sociolinguistic perspective in pronunciation teaching (Levis and Moyer; Deterding and Lewis; O'Neal). While other scholars have

focused instead on methodology and student's preferences in the ELF context (Burri; Saito; Nguyen). We can see how ELF is now at the center of pronunciation teaching research.

2.3. GRAPHO-PHONEMICS

Grapho-phonemics, in comparison, receives little attention. In fact, for Wijk (11), the development and teaching of grapho-phonemic rules should not be enforced in foreigners until they have acquired a large basic vocabulary. He saw no sense in learning the rules systemically having low proficiency of the language, as learning each word individually would work better. Learners would then naturally build on analogies and develop rules at their own pace.

Within the field of grapho-phonemics and their teachability, the focus has stayed for the most part in reading and spelling instruction for L1 children. However, the insights reached by some of these studies (Eddington et al.; Treiman et al.; Fry) can help with L2 pronunciation teaching. Fry's article—itself a summary and simplification of a previous study done by Hanna et al. in 1966— presents a classification of phoneme-grapheme correspondences. He gives the most common spellings for each vocalic sound found in a corpus of over 17,000 words (88). This is an example of how research tends to look from the perspective of spelling, i.e. how to spell each sound. Whereas an L2 learner would be more concerned with how to pronounce each letter.

We would argue that grapho-phonemics could be implemented in the understanding and teaching of pronunciation in ELF. As we mentioned previously, the English writing system is for the most part consistent through the globe (Hayes et al. 2005, 8), this would allow for a simplicity consistent with the ideas proposed by the LFC. From Jenkins in 2000 to Saito in 2019, the need of segmental accuracy as a basis for pronunciation teaching has been defended. Among others, the LFC calls for the need of "a clear distinction between long and short vowels" (Deterding and Lewis 4).

Applicable to the LFC and with a grapho-phonemic perspective, we can find Cámara-Arenas (2018). This study consists on measuring the level of representativeness of 10 pronunciation rules which predict the stressed vowel sound in different grapho-phonemic contexts. The prediction is based on the free and checked vowel contrast which Cámara-Arenas refers to as the nominal and alternative versions of the vowel. The rules would then help L2 speakers with the distinction of long and short vowels as stated in the LFC, providing a new tool in pronunciation teaching (as seen in figure 1).

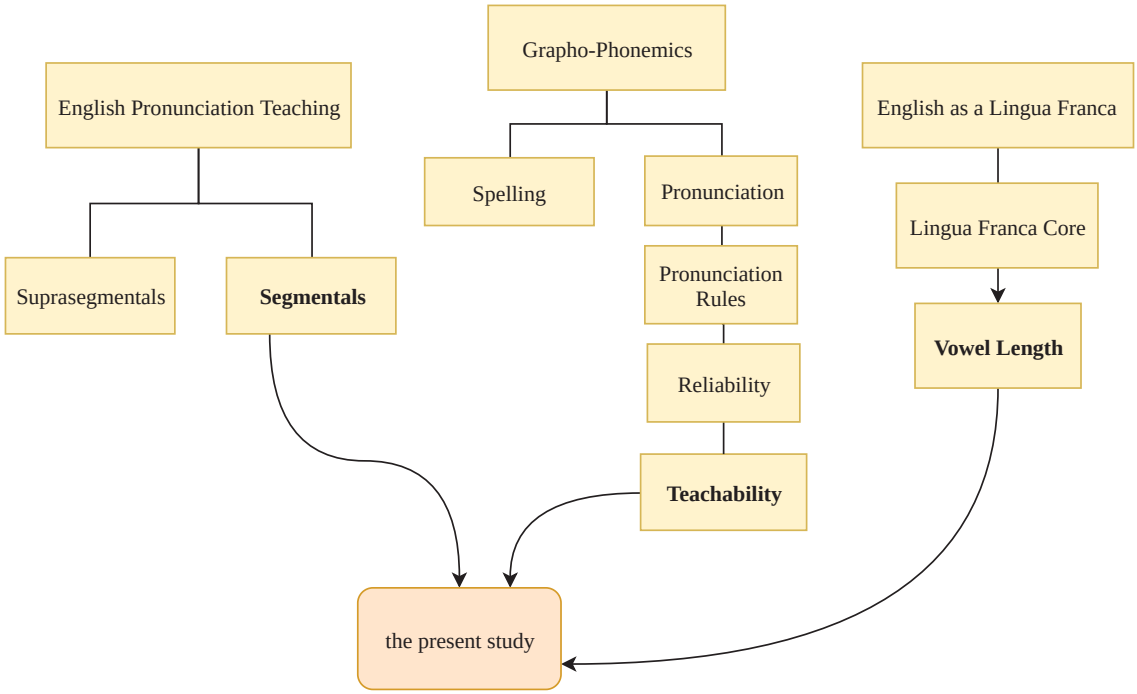

FIGURE 1: INFLUENCES IN THE STUDY

2.3.1. THE SPELLING SYSTEM

The English spelling system has been described in different ways. We will first have to understand that it is a "fundamentally alphabetic" system (Venezky 4), "as opposed to syllabic or logographic", but even if this consideration is simple enough, the notion of fundamentally is key: there are many aspects of English orthography that are more than purely alphabetic. This is what Venezky refers to as "the symbol-sound mismatch" (4).

This mismatch results in different possibilities. One could find issue with capital letters, apostrophes or simply asymmetry. As Venezky (4) states: capital letters "signal

word type or sentence position" while the apostrophe can signal "possession, contraction and elision". And, in the case of asymmetry, we can find one letter translating for two sounds or even into no sound, instead of the expected one letter equals one sound correspondence. A more complex issue is how a set of letters might translate to different sounds according to its place in the word, this would require for very clear context-related rules, as we will see. Venezky explores this idea of asymmetry by developing a number of properties and a set of seven principles that rule over English orthography.

Cummings contests the tag of "morphophonemic" previously in use. A morphophonemic system would be one that links orthographical units to not only phonemes "but also to morphemes, and therefore to grammatical and semantic units" (xxv). As Cummings develops the idea of the system being autonomous and self-governing, working in relation with other systems, such as morphology or syntax. The idea of interdependence is tied, along with the systemic demands, to other demands that are responsible for the nature of the system: Phonetic, Semantic and Etymological. He develops each in simple terms: "a given sound should be spelled consistently from word to word", "a unit of semantic content [should] be spelled consistently" and "a word [should] be spelled so as to reflect its etymological source" (12).

These three demands are in a constant state of balance in which they must also be compensated by the language's needs for standardization and ease of use. In fact, more often than not the etymological demand preserves "the user's notion" of the word's origin (12), giving weight to the system's reliance on conventionality.

It should be noted that English is a language with great flexibility in its acceptance of foreign words, "foreign spellings have been allowed to coexist with native ones" (Venezky 1999, 7) and the different retention patterns for spelling along time have also muddled the task of consistent rule making. With the formulation of the "principle of preservation of etymological traceability" (Cámara-Arenas 2010, 78), we have a link between Cumming's notions of origin and Venezky's visual identity. This principle not

only favors the etymological demand [1][4] but also the semantic demand [2] (Cummings), as it explains the need for morpheme reliability.

[1] the first <a> in *favorable*, according to standard pronunciation rules should be pronounced as /æ/ but instead is /eɪ/ to preserve its relationship to *favor*

[2] *boys* is not spelled as '*boyz' to keep the identity of the plural morpheme despite the sound being voiced

From Cámara-Arenas' article of 2018, we can glimpse the reach of this principle of etymological traceability when carrying out further examination on some of those words which do not follow the established pronunciation rules. However, etymology can only take the explanations so far and it would require a cultural knowledge that the L2 speaker is not required to have. The irregular pronunciation of *genre* would be easy to remember if one knows French, for example. The exceptions to the general-systemic rules would then need the application of domain-specific rules that would clarify the perceived irregularity.

3. THEORETICAL BACKGROUND

Grapho-phonemics deals with the link between grapheme and phonemes. These are the essential units of two different systems; the first based on letters and the other based on sound. Depending on the perspective through which we look at grapho-phonemics, we can focus on spelling or on pronunciation. If one were to describe a language for how close the equivalence between these two units is, we would speak of shallow and deep orthographies. For instance, the International Phonetic Alphabet would be the extreme example of a shallow system with univocal letter to sound correlation. Deep orthographies, such as that of English, involve more complex relationships which, according to certain scholars, lead to faulty rules, too many exceptions and recurrent claims for reform.

Through the process of acquisition, L1 children learn very early the patterns of their language both for spelling and pronunciation. L1 research is however focused for the

[4] Examples [1] and [2] are taken from Cámara-Arenas (2018) and Venezky (1999) respectively.

most part in reading and spelling. There has been little interest in the development of grapho-phonemic rules, defined as "translation rules for converting letters into sounds" (Castles et al. 98) with the perspective of L2 pronunciation.

Castles et al. explore the differences in the lexical and sublexical routes when children process new or unfamiliar words. Grapho-phonemics comes into play in the latter one, "which involves applying subword level grapheme-phoneme conversion rules" (99). For Spanish, having a shallow orthography, a learner would not need to access complex conversion rules. For example, when seeing a word like *agua* for the first time, an L2 Spanish speaker would process each grapheme individually to reach its corresponding phoneme.

For English, the sublexical route would require of a more complex thinking process. An L2 reader seeing *water* for the first time would need to be aware of many different grapho-phonemic rules. For example, they would need to know that the stress in this word falls on the grapheme <a>. Since it is followed by only one medial consonant, which is <t>, the general-systemic rule applicable to this type of word would be PR3.2 (Cámara-Arenas 2018).

According to this rule, the stressed <a> would be pronounced in the nominal version of that vowel: /eɪ/. However, because it starts with <w>, *water* is an irregular word and it is ruled by a domain-specific rule which only works for <a>. <w>, in this case, is a graphonemic indicator. These are "letters which may perform the function of indicating how other letters, especially vowel-letters, have to be pronounced" (Cámara-Arenas 2008, 91).

This domain-specific rule is pre-nuclear, which means that we have to look at the onset of the syllable. The onset of a syllable is what precedes the vowel. So <w> preceding a stressed <a> tells us that the vowel will be pronounced as /ɔ:/. If the L2 is aware of that then the word will be pronounced correctly. Pre-nuclear domain-specific rules are often subjected to other post-nuclear rules, e.g. the case of *wasp,* "the satisfaction of onset specifications by itself does not allow vocalic grapheme interpretation" (Cámara-Arenas 2008, 97).

Post-nuclear rules are much more efficient and exhaustive than pre-nuclear, due to a stronger relation between a vowel and its coda. The coda is what follows the vowel, the combination of these two "belong to the rime constituent of the syllable" (Fudge qtd. in Treiman et al. 449). This strong relation can be attributed to how "the spelling system has more associations between vowels and codas than between vowels and onsets" (Treiman et al. 463).

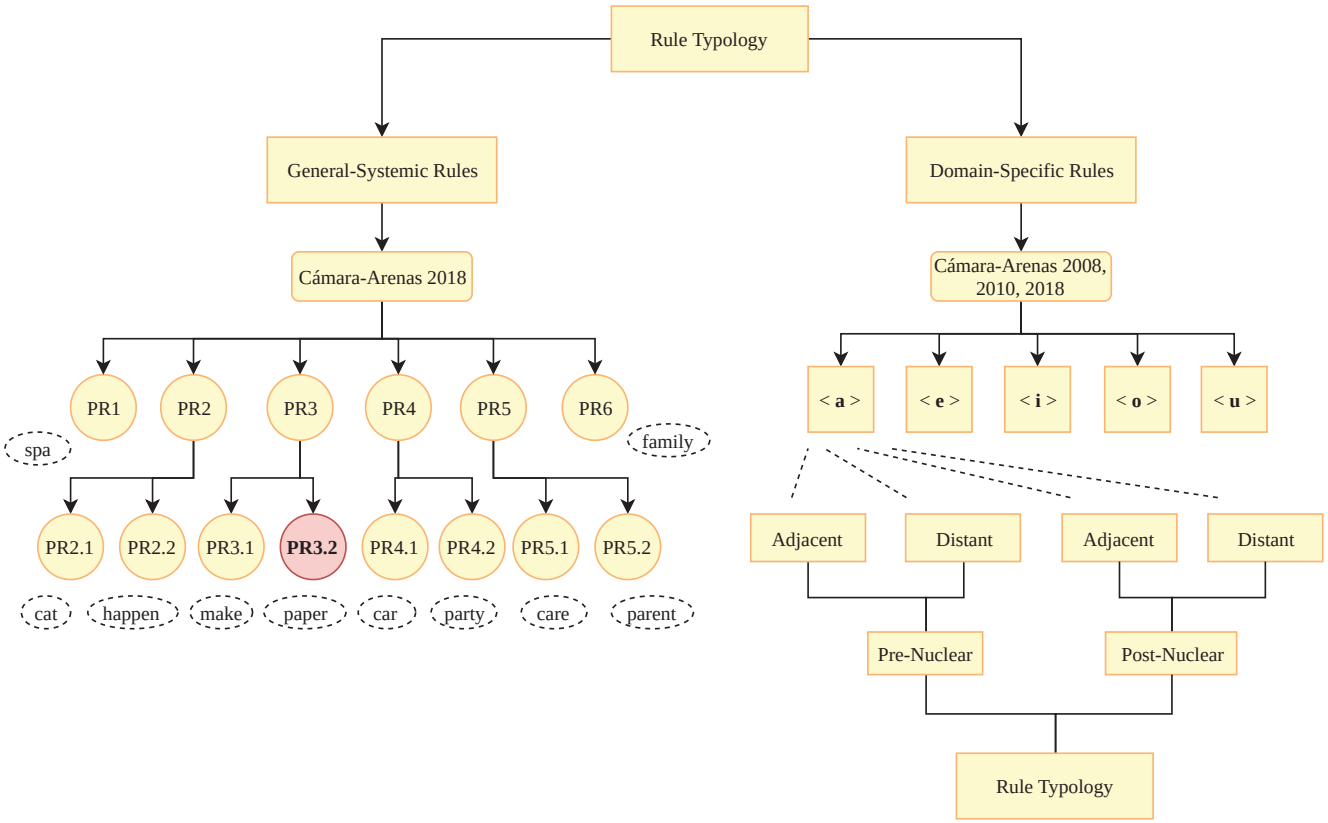

FIGURE 2: RULE TYPOLOGY

At the same time, rules can also be classified according to how exact is the relation between the graphonemic indicators to the stressed vowel grapheme. This would give us adjacent and distant domain-specific rules. Adjacent rules would be applied only when the graphonemic indicators in immediate preceding or following position to the vowel. Distant rules would be for contexts which are not directly surrounding the vocalic grapheme.

Consonant graphemes found between two vocalic ones belonging to two different syllables are called *medial* (the medial consonant of the word 'medial', for example,

would be <d>. Depending on the characteristics of this consonant the stressed vowel will be affected in different ways (Treiman et al. 464), adjacent post-nuclear domain-specific rules would be applied in this case [3].

[3] *exit, lexis, exile* obey adjacent post-nuclear domain-specific rule <e>+<x>+V: /ɜ/

Distant rules, for example, would apply for words ending in specific suffixes for which it does not matter how many graphemes separate the stressed vocalic grapheme from the indicator [4].

[4] The false proparoxytone[5] rule is a post-nuclear distant domain-specific rule which can be applied to <e> in words like *d**e**cade* or *pr**e**sence.*

A classification of rule typology (see figure 2) can provide a clearer idea of how complex it can be for an L2 speaker of English to select between all possible pronunciations of a stressed vocalic unigraph. The sublexical route can lead the speaker to any of the 10 general-systemic rules, however, in the case of general-systemic rules failing, the L2 would have to apply a domain-specific rule (which in some cases also have exceptions).

The English paroxytone, examples [5], [6], [7] and [8], is one of the most common structures of the English lexicon. In Cámara-Arenas' (2018) sample of 5000 words, the paroxytones amount to 42.2% of the total. They are divided in two different rules: PR2.2. predicts the pronunciation of /æ/ for a stressed <a> in a word with two or more consonants in the medial position. While PR3.2. predicts a pronunciation of /eɪ/ for the <a> in a stressed position followed by only one medial[6].

[5] regular PR2.2. word: *h**a**ppen* (/æ/)

[6] irregular PR2.2. word: *ch**a**mber* (/eɪ/)

[7] regular PR3.2. word: *p**a**per* (/eɪ/)

[8] irregular PR3.2. word: *p**a**nel* (/æ/)

---

[5] Proparoxytones are those words in which the stress falls in the antepenultimate syllable, e.g. *par**a**lysis* or *pr**e**sident.*

[6] We have taken examples from the <a> domain, but all general-systemic rules are applicable to the five domains: <a>, <e>, <i/y>, <o> and <u>.

A vowel grapheme followed by a double consonant will most probably be pronounced with the short or alternative version of that vowel. While a vowel grapheme closed by only one consonant will be pronounced with the long or nominal version. This is consistent with more recent findings on how children are already aware of both the length of a vocalic phoneme or the number of vocalic graphemes when spelling the following consonants (Hayes et al. 2005).

But, besides helping with L1 spelling, the development of these rules for the purposes of pronunciation teaching could be very useful for L2 students. In Cámara-Arenas 2018, a set of ten rules for stressed unigraphs in different contexts has a general reliability of 87%. PR2.2. has 92% accuracy, while PR3.2. only has a predictability of 58% for all vocalic unigraphs.

Being the third most common structure in the database, the 3.2. structure represented 18% of the total 5000 words. With such a high number or irregular words, the author questions the validity of rule 3.2. altogether. However, many of these exceptions might "actually follow other easy and reliable domain specific rules" (210).

PR3.2. predicts when a stressed vocalic unigraph will be pronounced with the nominal version of the vowel (also known as "free pronunciation", Wijk 36). Thus, a stressed <a> will be /eɪ/; stressed <e>, /i:/; stressed <i> or <y>, /aɪ/, stressed <o> would be /oʊ/ and, finally, stressed <u> would be either /u:/ or /ju:/. This rule works for words like *paper, Peter, final, open* and *human*[7] but it fails to predict the pronunciation of others like *publi*c or *cover*. This idea falls into the so-called VCC/VCV contrast (Cummings 96), VCV is the first of the three strings included in PR3.2.

In an analysis of native speaker syllabification, Eddington et al. had 80% of their participants in agreement on where to place syllable boundaries for only half of a set of disyllabic words with only one medial (1: 60). The fact that L1 speakers could not be sure of syllable split for this type of words (PR3.2. structure) could be relevant to how irregular its pronunciation is. Especially when considering the notion of sonority.

[7] Examples taken from Cámara-Arenas 2018.

"The more sonorous the medial consonant or first consonant in a CC cluster, the more likely it is to be placed into the coda of the first syllable" (Eddington 2: 90). If we take this into consideration a consonant placed at the coda of a vowel might trigger the alternative or short pronunciation of the vowel. This agrees with PR2.2.'s prediction of alternative pronunciation (in a VCC situation) and could also help with singleton medials in the VCV string of PR3.2. if they are sonorous.

The other two strings considered in PR3.2. are part of what Cummings calls the "holdout" (105) of the VCC pattern: this happens with the case of the two liquid consonants, when we substitute the second consonant with <r> and <l> we obtain PR3.2.b (examples [9] and [10]) and PR3.2.c (exammples [11] and [12])respectively. Cummings limits the string to "VCle#"[8], considering only those instances in which the stressed syllable is followed by a "word-final *e*" (105), in our research we implemented all possibilities as seen in Cámara-Arenas (2018), who enunciates the string as "…vc<l>v…" (204). Lastly, the "VCrV[9]" pattern as called by Cummings who already expressed it as having "apparent holdouts" (106).

[9] regular PR3.2.b word: *sacred* (/eɪ/)

[10] irregular PR3.2.b word: *fabric* (/æ/)

[11] regular PR3.2.c word: *able* (/eɪ/)

[12] irregular PR3.2.c word: *establish* (/æ/)

Our aim in this study is to see how valid the application of domain-specific rules is in the justification of rule exceptions for each vocalic unigraph. This should be done while staying within the teachability approach. An exhaustive exploration of these domain-specific rules would probably not be advisable or sensible. Considering previous studies, there is hope in the application of these rules to the exceptions of PR3.2 (Cámara-Arenas 2018, 213).

---

[8] According to our method this formula would be: vc<l><e>#.

[9] vc<r>v according to our system.

4. METHODOLOGY

4.1. ENGLISH VARIETY

Wijk states that the pronunciation differences between cultivated varieties of American and British English are not "reflected in the spelling of the language" (12). The minor spelling differences (such as '-our' and '-or', '-re' and '-er', '-ce' and '-se' or '-ll-' and '-l-') do not have any phonetic significance. While the three main aspects of differentiation are those pertaining the vowels <a> and <o> in certain contexts and the approximant <r> in "final and preconsonantal position" (12). None of which matters much for our study. Since these dissimilarities between British and American are for the main part regular, our ideas in this study could be as easily applied any variety, however we have worked with a list of words taken from COCA, and as such, belonging to the General American standard variety of English.

4.2. DATA

We worked with Davies' list of the most common 5000 words[10] used in American English as found in the COCA corpus. We needed to cross-reference Davies' list with a dictionary in order to get the phonetic transcription of each word. So, we selected the Carnegie, for two reasons: the phonological representation of each word also followed the American variety and it was done following the ARPAbet convention. ARPAbet is a UTF-8 based machine-readable phonetic alphabet, and therefore amenable to automatic processing.

After a brief analysis of the data we noticed some issues: Davies' list is tagged for grammatical category. We would have words like *contest*, once tagged as "noun" and then again tagged as "verb". Grapho-phonemically, this would be considered as only one word, regardless of function. This resulted in our list having as many as 1256 duplicates. This issue posed a problem. In some instances, a word is marked for function in where the stress is placed. Our solution was to consider only each word once without tagging

[10] http://www.wordfrequency.info

for syntax. Our script only took the first entry of each word in the dictionary, taking the first option as dictated by Carnegie.

Davies' list also had words which were not found in the dictionary such *as 't, ie, mm-hmm, and/or* or *self-esteem*, we also cleared these out.

We had to filter words that had undergone morphological processes such as Compounding, Derivation and Verbal Inflection. In accordance with Cummings' System Demands (12) and Cámara-Arenas' principle of preservation of the etymological traceability (2010, 78). Many of the words tagged as exceptions seemed to obey the semantic and etymological demands of a spelling system who at times is not over concerned with the representation of sounds.

For example, our program classified words like *therefore, somehow, something, somewhere, someday* and *sometime*, *online* and *sunlight* as paroxytones. We can see how in all instances; these actually obey pronunciation rules applied to monosyllabic words. Similarly, *careful, badly, severely, lover* and *lovely* are derived from monosyllabic words and as such, should be ruled by different pronunciation rules. *Business,* as well, was taken out and only *busy* was considered. Among our words, we also found *living, given, thanksgiving, during* and *coming*, which had been classified as paroxytones but are pronounced as the uninflected version of the verb.

4.3. PROCESS

While Cámara-Arenas classified the data manually, we decided to proceed automatically. We worked with Python 3 for the two main parts of the process: the preparation of the data and the selection of PR3.2 words. That is, we developed a way to automatically filter and classify our data. The program could then be given any number of words and would return whether they complied or not with pronunciation rule 3.2.

4.3.1. PREPARATION

We set down an automated substitution protocol: such as 'v' working as a substitute for vowels, while 'c' worked for the consonants. Evidently, the matter was not this simple, we decided to create a group of so-called 'exceptions' that we wanted to keep separate from other characters in the formula: l, r, w, y, h, e. In some cases, these graphemes had a double functionality, e.g. <y> and <w>. <y> is pronounced as a consonant in *yet* but as

a vowel in *cycle*. In other cases, it was because they often act as graphonemic indicators (Cámara-Arenas 2008, 91). For instance, we separated <e> from the other vocalic graphemes, since its presence in word-final position often obeys grapho-phonemic reasons, i. e. selecting the nominal version for the pronunciation of the stressed vowel in *cane*. <h> was also kept apart for its frequency in clusters such as <gh>, <ch>, <ght> which sometimes affect preceding vowel sounds. And, finally, <l> and <r> had to be isolated in order to look for the specific contexts required in PR3.2.b and PR3.2.c.

Likewise, since the literature can indicate certain contexts where these graphemes do not require any further examination, we substituted the <y> or the <w> in final positions for a vowel. Similarly, some of the other consonants could give us problems in some circumstances, so we also wrote substitution rules to, for example, classify the cluster '<gu>+any vowel which was not <e>' as 'ccv', instead of 'cvv'.

Furthermore, we cross-referenced the word list with the Carnegie dictionary. And then we separated the phonetic transcription and the stress pattern. We needed to be able to deal separately with stress. A word like *cover* (number 559 in the list) would have a Carnegie transcription of K AH1 V ER0[11]. After the separation we had K AH V ER and a stress pattern of [-1, 1, -1, 0]. Once data preparation concluded, we had all the necessary information in five columns (figure 3): index (which indicated the order in which the

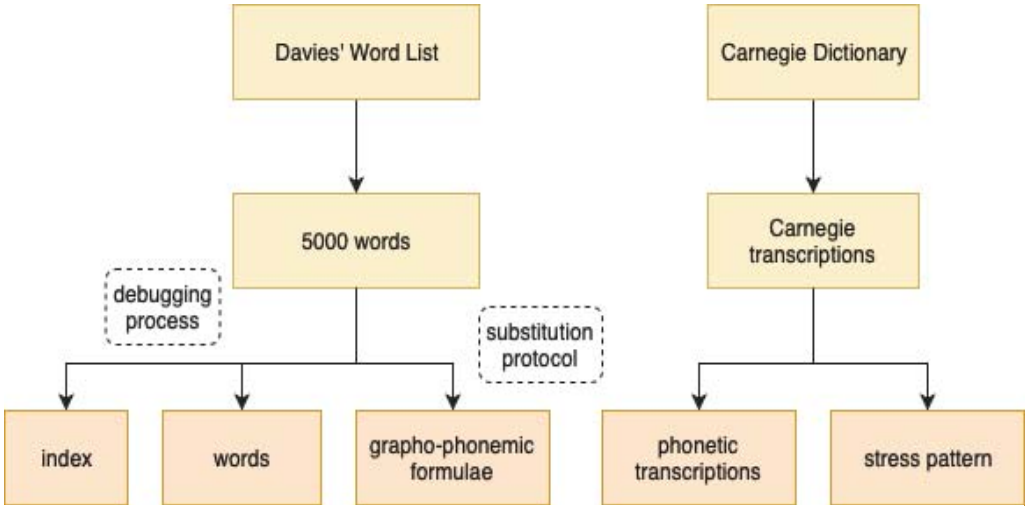

FIGURE 3: DATA TREATMENT PROCESS

[11] /'kʌvər/ in standard IPA.

word appeared in Davies' list), the word, its grapho-phonemic formula (developed by means of the substitution protocol), its phonetic transcription and its stress pattern (both taken from Carnegie).

4.3.2. SELECTION

The process of selection was itself divided into four steps. To focus only on PR3.2, we needed the program to get only the paroxytones from the list. This issue comprised the first and second steps: firstly, we needed to clear out all monosyllabic words and, secondly, we needed to keep only those for which the primary stress fell on the second to last syllable. Once we had selected the paroxytones from the list we needed to keep only those to which PR3.2. could be applied (third step). Lastly, we selected only the words whose stressed vowel was not pronounced according to its nominal version (fourth step).

We have taken a set of seven words included in Davies' list in order to better show the different phases of the process as words are kept or discarded, the words are: *abroad, age, client, cost, ever, fever* and *matter.*

FIRST STEP

We marked as "Valid" those words for which the vowel and/or digraph count was superior or equal to two. Vocalic digraphs had to be classified so a word like *feat* would

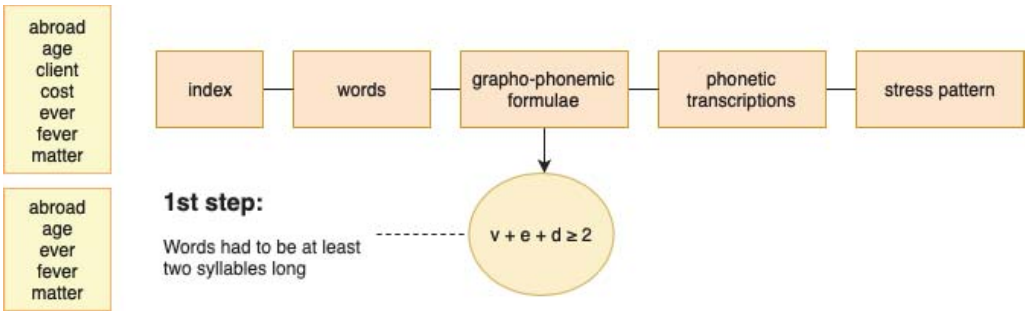

FIGURE 4: FIRST STEP

not be considered as disyllabic. Since the code was written with the specific purpose of selecting words for the PR3.2, we simplified some issues: for instance we did not consider the possibilities of hiatus in our tagging of vocalic digraphs since none of the strings where PR3.2 is applicable contain a v.v combination, consequently words like *lion*, for example, were considered monosyllabic even if they are not.

The program looked for three different things: the vowels as specified previously, the digraphs (considering 'vv', 'v<e>', '<e>v' and '<e><e>' and replacing them for "d") and the isolated <e>. If these three factors counted superior or equal to two, the word was considered Valid, if it were inferior the Boolean gave a value of False and the word was discarded. In the example above (figure 4), we can see why the tagging of <e> as an exception was necessary: words like *age* were initially classified as disyllabic.

SECOND STEP

The issue of how to classify the paroxytones was problematic. We needed a way to assign the mark of stress to either the valid words or their formulae. This is why we required the isolated stress pattern from the phonetic transcription given by Carnegie.

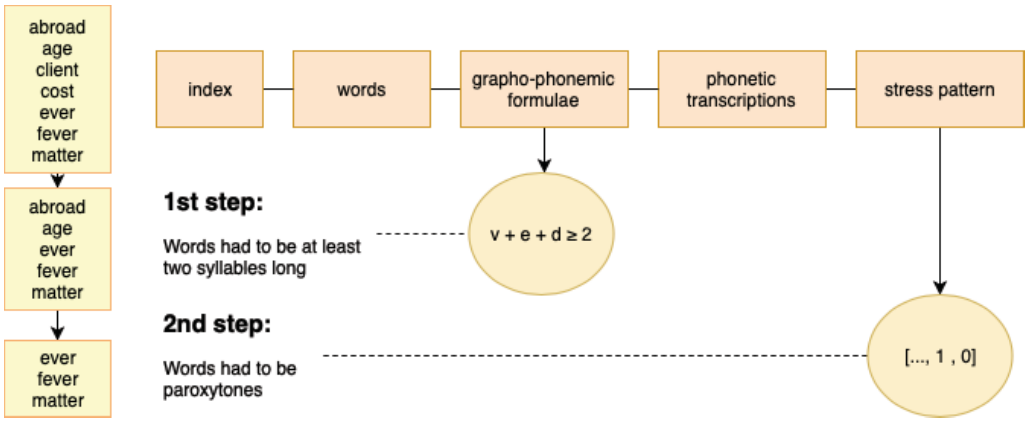

FIGURE 5: STEPS 1 TO 2

A word like *fever* for example, gave us a '-1, 1, -1, 0' pattern. We gave the negative value to sounds with no stress and isolated all other marks from the vowels. Then, we asked only for those words that, counting from the end, had a 0 or a 2 first, followed by a 1, the negative values were ignored. *abroad* returned a stress pattern of [0, -1, -1, 1, -1], note that this was done to the phonetic transcription and not the grapho-phonemic formula.

THIRD STEP

Once we had the list of paroxytone words we wrote into the program the three sets of strings (we called them "clusters") for which PR3.2. is applicable: a) vcv, b) vc<r>v and c) vc<l>v (with the corresponding variants for the isolated <e>), Python would then find the clusters in the grapho-phonemic formulae, effectively giving out a list of the paroxytone words whose pronunciation should be predicted by PR3.2.

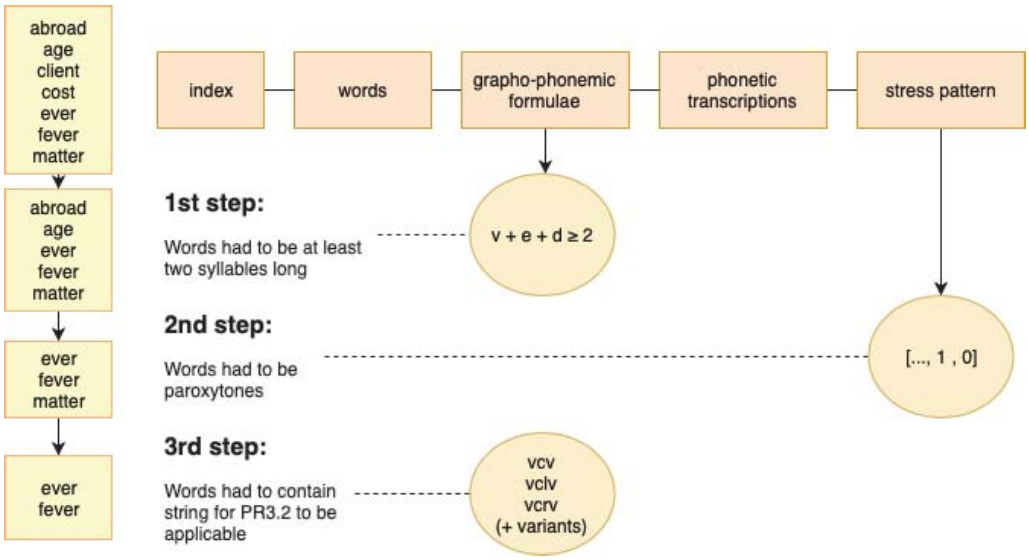

FIGURE 6: STEPS 1 TO 3

FOURTH STEP

Lastly, we had the set of words to which PR3.2. was applicable. Since the focus of our study was on the exceptions to this rule, we looked for the expected pronunciation. We could do this directly by taking those words that had the nominal version of the vowel in a stressed position, those were regular PR3.2. words like *fever*.

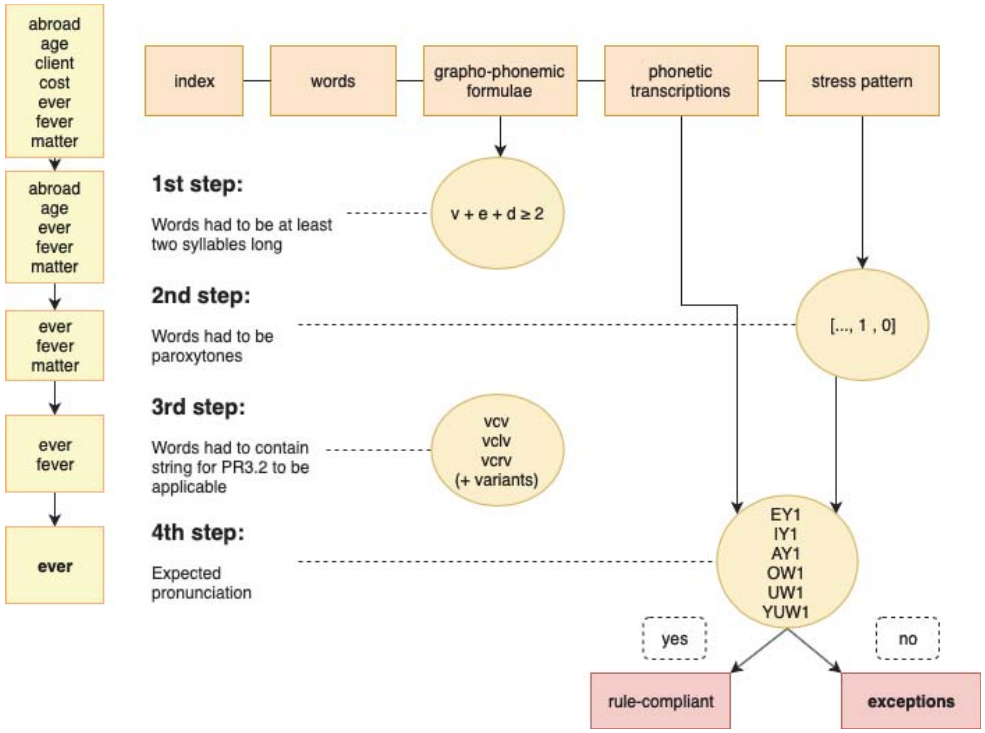

FIGURE 7: STEPS 1 TO 4

We were then left with the exceptions to PR3.2., words like *ever* in figure 7. Through the whole process we had taken out everything but the irregular paroxytones from the 5000 words in Davies' list, but there was one final issue with our results that needed further sifting.

We had looked for the string within a paroxytone, but we had no way of selecting the location of the string and isolating it to its place around the stressed vowel. Because of this we had to filter the results one last time, taking out words like *democr**a**tic*. The word had a paroxytone structure but the '**v**c<r>v' was placed around an unstressed vowel[12].

5. RESULTS

5.1. GENERAL

After the final process of debugging we were left with 219 irregular words according to pronunciation rule 3.2. The words were unevenly distributed along the three strings in which PR3.2. is applicable. PR3.2.a (vcv) amounted for 198 of the irregular words, while PR3.2.b (vc<r>v) and PR3.2.c (vc<l>v) had only three and eight respectively (see table 1 below).

	Total	**PR3.2.a**	**PR3.2.b**	**PR3.2.c**
A	56	54	1	1
E	34	33	1	0
I	70	69	0	1
O	49	37	1	1
U	10	5	0	5
	219	198	3	8

TABLE 1: DISTRIBUTION OF EXCEPTIONS

Also relevant to our analysis is the distribution of irregular words among the five different stressed unigraphs. Exceptions for <u> represented only 4% of the total, while the irregular pronunciations of <i> amounted for almost 32% of the whole.

[12] The stressed vowel is marked in bold type.

Results were also classified in terms of the phonetic realization: we had as many as seven different possibilities to the rule. PR3.2. predicts the nominal version of the pronunciation for the unigraph. In some cases (<e> and <i>), irregular words had only one other alternative pronunciation[13] (/ɜ:/ and /ɪ/ respectively). However, for <a>, <o> and <u> there were as many as four or three possibilities (table 2 below).

Count	Total	ɑ:	æ	ɔ:	ʌ	ɛ	ɪ	ʊ
A	56	2	51	1	0	0	0	0
E	34	1	0	0	0	33	0	0
I	70	0	0	0	0	0	70	0
O	49	36	0	1	11	0	0	1
U	10	0	0	0	8	0	1	1
	219	39	51	2	19	33	71	2

TABLE 2: DISTRIBUTION OF ALTERNATIVE PRONUNCIATIONS

5.1.1. WORD CHANGE

There were four instances in which words from our selection were considered by Carnegie as paroxytones (stress placed in second to last position), while other dictionaries would classify them as proparoxytones (stress placement in third to last position). This can be due to the elision of a schwa in the intermediate position. The pronunciation of the stressed vowel in these four words would then be predicted by PR6 (Cámara-Arenas 2018, 204). For example, *opera* was considered to be disyllabic: /'ɑ:prʌ/ instead of /'ɑpərə/[14]. Three of them are regular according to PR6, which rules over proparoxytones (*elaborate*, *opera* and *chocolate*) while one is not (*recovery*). These words are accounted for in the first column from the right in table 3.

	Domain-Specific Rules											
	Pre-Nuclear	Post-Nuclear										
	Adjacent			Distant						others		
	<w>	medial <v>	medial <x>	V+C+ <i-...>/#	V+C+ vv...	false prop	V + C + <on(d)/or(d)>	open syllable	ard	ent	al	elision of schwa
A	2	1	0	27	3	9	0	1	1	2	0	1
E	0	6	2	8	2	5	2	0	0	1	2	0
I	0	4	0	15	39	6	0	0	0	0	0	0
O	1	5	0	13	1	3	0	0	0	0	0	3
U	0	0	0	4	0	0	0	0	0	0	0	0

TABLE 3: DOMAIN-SPECIFIC RULES APPLIED IN ALL 5 DOMAINS

[13] Except for /ɑ:/ being the pronunciation of the stressed <e> in *genre*.

[14] Found in J.C. Wells' Longman dictionary.

Furthermore, in table 3 above, we can see the complete numbers for the different domain-specific rules applicable to our set of PR3.2. exceptions. These are not all domain-specific rules in the literature (Cámara-Arenas 2010), but the ones we could apply to the words in Davies' list. Note that the rules have been classified according to their typology (as developed in the Theoretical Background section of this study) and to the different domains where they apply. The cells filled in light orange are exceptions to the rules, while the cells in green correspond to examples listed in the literature (Cámara-Arenas 2010). Cells with no color fill represent the figures of a tentative classification, this is especially significant considering <o>, for which there are only two domain-specific rules (Cámara Arenas 2010, 162).

We will go over the casuistic and application of the two most effective rules:

	V+C+vv…						
	ious	**ion**	**ian**	**iar**	**ial**	**ient**	**uV**
	4	27	6	1	4	2	2
A	0	1	1	0	0	0	2
E	1	0	0	0	1	0	0
I	3	25	5	1	3	2	0
O	0	1	0	0	0	0	0
U	0	0	0	0	0	0	0

TABLE 4: RULE V+C+VV...

Seen above (table 4) is the second most applicable rule, with 46 cases. This rule is supported mostly in its application to the domain of <i> in the literature (Cámara-Arenas 2010, 132), but the case of *statue* and *value* and of *onion* are also supported (66; 160). *Italian, companion* and *special* are exceptions to the rule in their respective domains (64; 98). While *precious* has not been registered yet.

	V+C+ <i-…>#				
	ic	**id**	**ish**	**it**	**in**
	41	4	6	9	2
A	17	2	3	1	2
E	6	0	0	2	0
I	6	1	3	5	0
O	10	1	0	2	0
U	2	0	2	0	0

TABLE 5: RULE V+C+<I...>#

This rule is the most effective for our word list, it is applicable to four of the five domains with a total of 67 cases. As seen in table 5, there are some instances in which we have found words for the domain of <u> that could fit into this rule. However, the rule is not applicable to a sufficient number of words to be reliable. We have *public* (/ˈpʌblɪk/) but *pubic* (/’pju:bɪk/), for example.

5.2. REALIZATION OF <A>

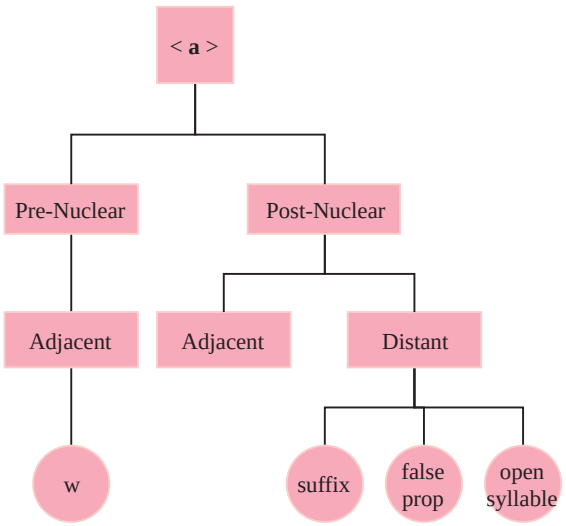

FIGURE 8: DOMAIN-SPECIFIC RULES APPLICABLE TO <A>

There are four domain-specific rules (see figure 8) applicable to the 56 exceptions of PR3.2. stressed <a>. We can justify as many as 40 of these words with domain-specific rules seen in Cámara-Arenas 2010 (64-66).

In other cases, the words are themselves exceptions to other domain-specific rules. For example, we have 4 words with a stressed <a> that end in a free syllable. The rule, also in Cámara-Arenas (2010, 74), <a> + C + V#, allows us to redeem *drama* with its pronunciation of /ɑ:/, and an alternative pronunciation of *banana*[15]. Exceptions to this rule, [16]however, are found in *any* and *many* (/ɛ:/).

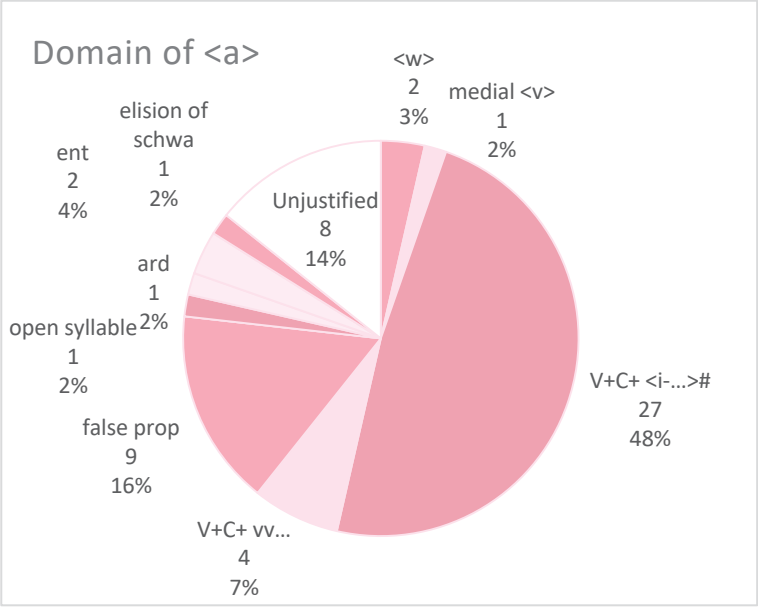

FIGURE 9: DISTRIBUTION OF EXCEPTIONS FOR <A>

In figure 9 above we can see how the final percentage of unjustified words for <a> falls to 17%, this would bring <a>'s reliability to a level on par with Cámara-Arenas' results (2018) for other rules.

[15] Although Carnegie gives the stressed <a> in *banana* a pronunciation of /æ/, we have found instances in which it is given /ɑ:/.

[16] For all pie charts: dark-filled sections are rules found in the literature, light ones are either exceptions to those rules or unregistered, this is consistent with the green/orange coding in tables 3-5.

5.3. REALIZATION OF <E>

Of the total 34 exceptions for stressed <e>, we can justify 23 with rules found in Cámara-Arenas 2010. We can apply 5 different rules to this domain, with different levels of success. Medial <v> (102) for example works with all words derived from *ever* as well as many others like *clever* or *level,* however it does not work for *even* or *fever*, which follow the general PR3.2. rule.

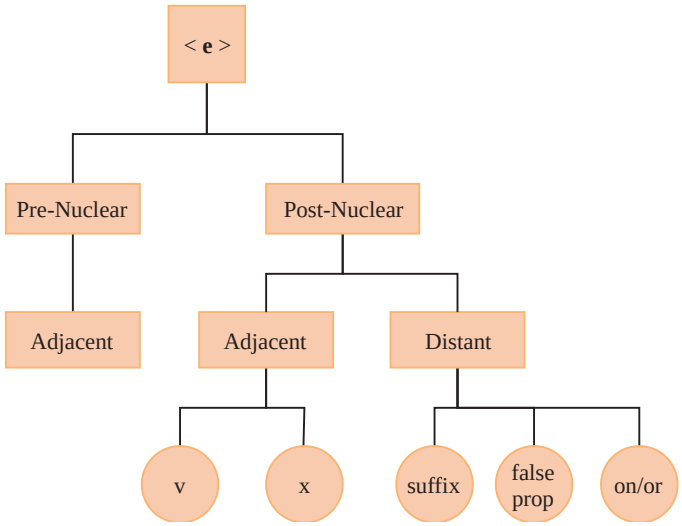

FIGURE 10: DOMAIN-SPECIFIC RULES APPLICABLE TO <E>

As seen in figure 10 above, <e> has a very specific rule that we have simplified as on/or, this rule is only useful for a small number of cases like *lemon* or *second.* This is an example of the reliability/teachability dilemma. Although the rule works well in this domain, we would not advise it being taught due to how little is its reach.

We have two other words already explained as exceptions to rule v+c+vv. We have grouped three more under different contexts: *present*[17] is in a tentative group with *patent* and *talent*, while *medal* and *metal* are registered as dubious in Cámara-Arenas 2010 when compared with *legal* or *penal* (98). See figure 11 for the overall distribution of words.

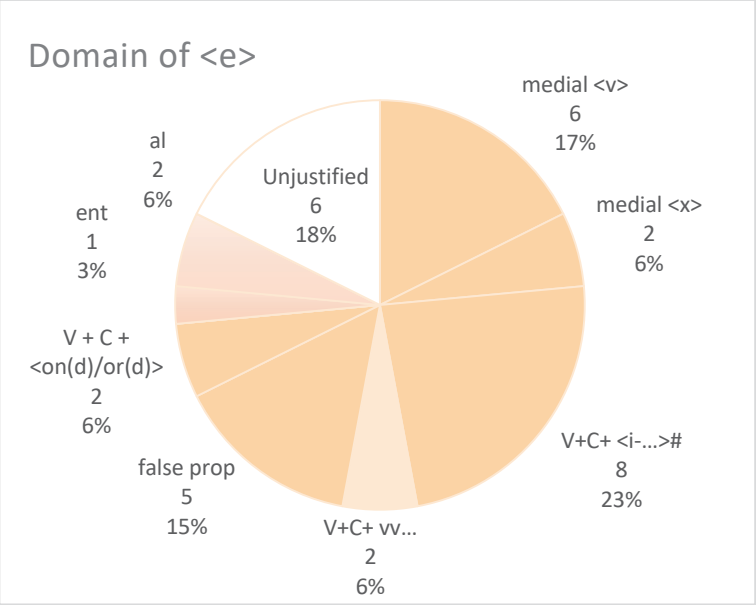

FIGURE 11: DISTRIBUTION OF EXCEPTIONS FOR <E>

5.4. REALIZATION OF <I>

Words with an <i> in stressed position took almost 32% of the total of PR3.2. exceptions, however, this is the most regular of all irregular domains with only one possible alternative pronunciation. The exceptions of <i> could be justified up to a 91% along just four different domain-specific rules (see figure 12). The only six words which we could not find any domain-specific rules for are: *city, sibling, widow, consider, prison* and *continue.*

[17] However, it cannot be paired with *recent* or *frequent.*

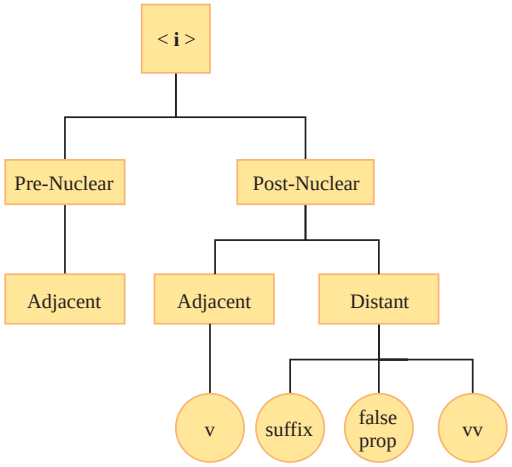

FIGURE 12: DOMAIN-SPECIFIC RULES APPLICABLE TO <I>

For the case of <i>, all exceptions justified via domain-specific rules were found in the literature (Cámara-Arenas 2010), see figure 13 below, for the overall distribution of words in the domain-specific rules applicable to <i>.

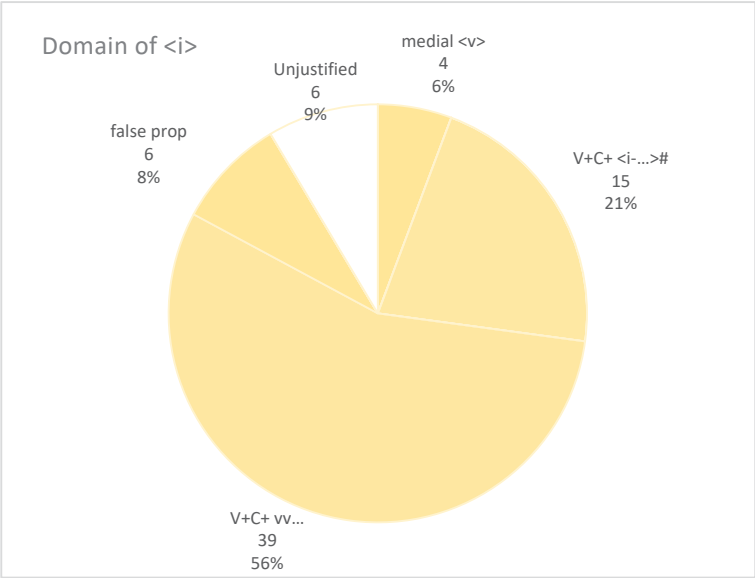

FIGURE 13: DISTRIBUTION OF EXCEPTIONS OF <I>

5.5. REALIZATION OF <O>

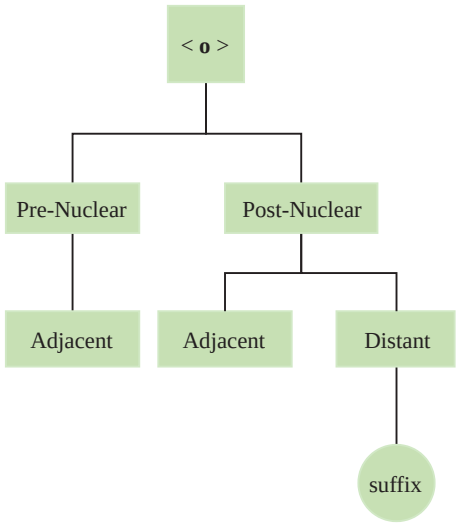

FIGURE 14: DOMAIN-SPECIFIC RULES APPLICABLE TO <O>

Among all five stressed vocalic unigraphs, <o> is the most problematic. We have found only one domain-specific rule applicable to the words in our list, the aforementioned v+c+<i…># rule (see figure 14 for its classification).

The percentage of final unjustified exceptions to PR3.2. is the highest among all five domains at 47% (see figure 15 below). And the rules applied are, in some cases, not supported by reliable evidence. Only 17 of the 49 total <o> exceptions are registered in Cámara-Arenas 2010 .

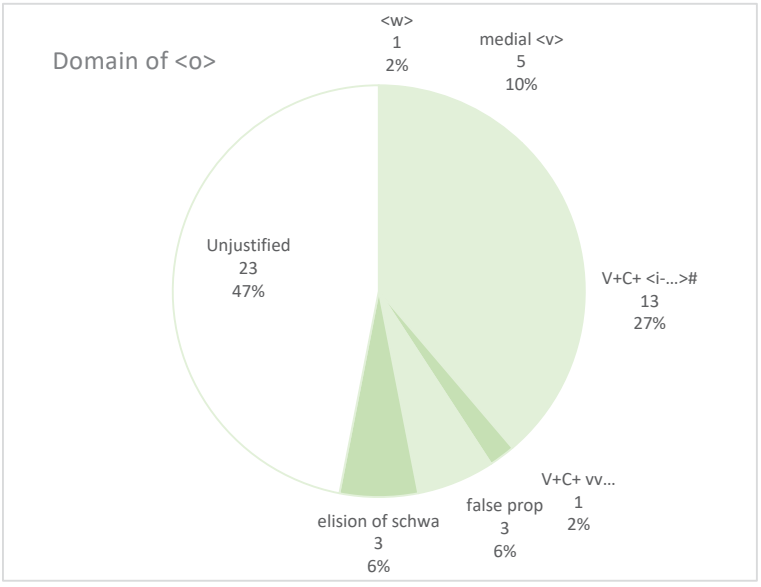

FIGURE 15: DISTRIBUTION OF EXCEPTIONS OF <O>

5.6. REALIZATION OF <U>

We have found no words among our exceptions with evidence in the literature (see figure 16). <u> has the fewest exceptions among all domains, but the exceptions are notably varied.

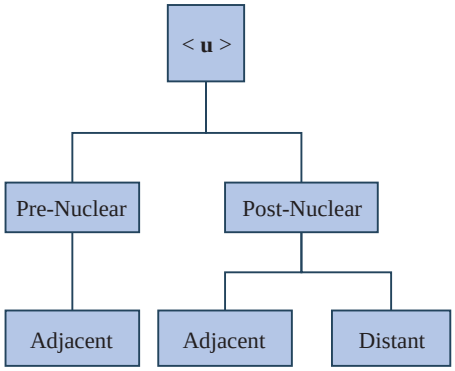

FIGURE 16: DOMAIN-SPECIFIC RULES APPLICABLE TO <U>

We have to point out *busy* as an exceptional pronunciation of /ɪ/, that is taken by all its variants (as seen with *business*). Another isolated case is the common *sugar* with its /ʊ/ being the only one occurrence of this phoneme for all the study. That leaves us with the more frequent /ʌ/ phoneme for an unstressed <u>, which we can find in *Muslim*, *study* and *suburb*. The latter probably easily attributed to being an abbreviation of the regular proparoxytone *suburban*. We can see in figure 17 the representation of words like *public, punish, republic* and *publish*, which were already accounted for previously.

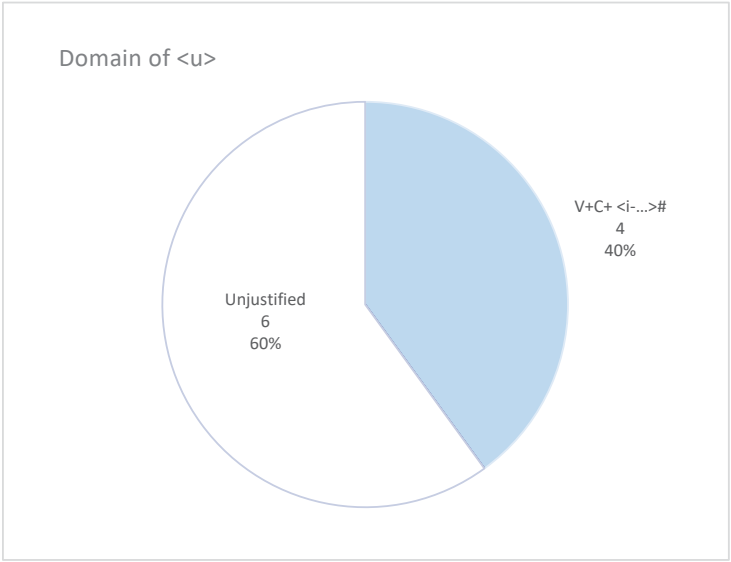

FIGURE 17: DISTRIBUTION OF EXCEPTIONS FOR <U>

6. DISCUSSION

	Cámara-Arenas 2018	
	Regular	**Iregular**
A	140	53
E	31	38
I	34	75
O	51	49
U	57	12
	313	227
	58%	42%

TABLE 6: CÁMARA-ARENAS 2018 RESULTS FOR PR3.2 REGULARITY

Following the results found in Cámara-Arenas (2018) (table 6), we can see that the accuracy of Pronunciation Rule 3.2. was of 57%, with 313 regular items of the 540 total. The 227 irregular words found in his article are close to our figure of 219, which is to be expected as we developed a different method and criteria. While we developed an automatic process of data selection and classification, the earlier article proceeded manually. Cámara-Arenas also selected a different dicionary for the phonetic representations, taking J. C. Wells' (Longman) dictionary, and we selected Carnegie for reasons explained in the Methodology.

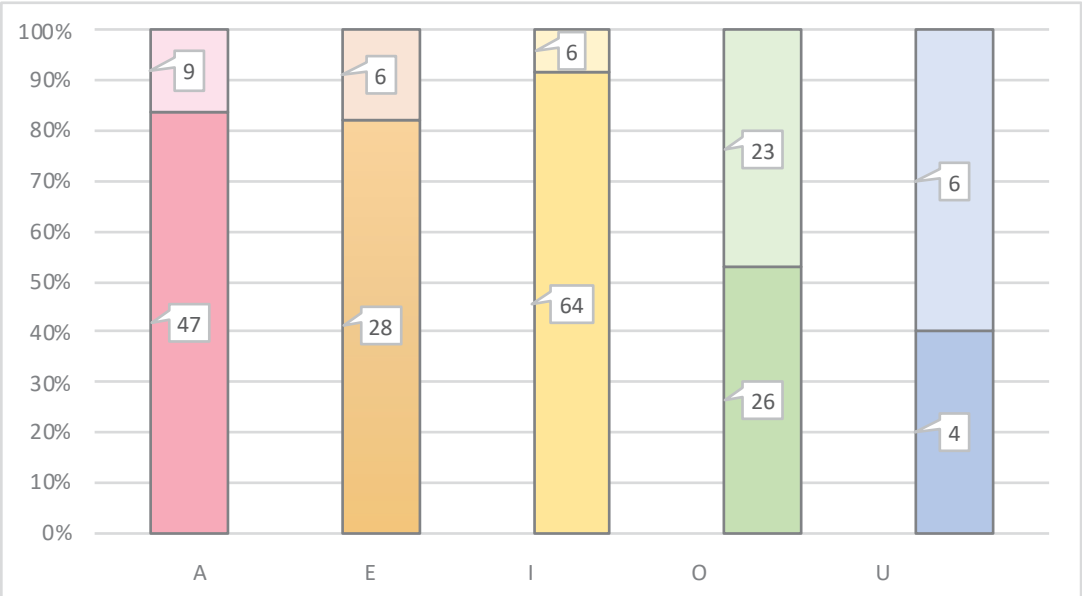

FIGURE 18: FINAL NUMBERS FOR ALL DOMAINS

We have redeemed these irregular words with the application of domain-specific rules, arriving to a final count of only 49 unjustified irregular words over all five domains. As seen in figure 18, domains <a>, <e> and <i> have achieved considerably high levels of regularity through the application of these domain-specific rules. While <o> maintains an average comparable to the general applicability of PR3.2, and the few cases in <u> cannot be redeemed with domain-specific rules. <o>, the third domain in total number of exceptions to PR3.2., does not obey the domain-specific rules that we have found in the literature to the degree that <a>, <e> and <i> do, bringing the general average of justified words lower. In total, the 170 justified words represent 78% of the exceptions found to PR3.2. (see table 7).

		PR3.2. Exceptions	
	Total	**Justified**	**Unjustified**
A	56	48	8
E	34	28	6
I	70	64	6
O	49	26	23
U	10	4	6
	219	170	49
		78%	22%

TABLE 7: JUSTIFIED AND UNJUSTIFIED EXCEPTIONS TO PR3.2.

Considering the reliability of PR3.2. words in general, the application of domain-specific rules in addition to the general-systemic would increase the general reliability of paroxytones with only one medial consonant. When adding our justified words to the number of regular words found in Cámara-Arenas 2018, words with a PR3.2. structure have a 89% total dependability (table 8 below for individual unigraph figures).

	Cámara-Arenas 2018			Regular & Justified	
	PR3.2. Total	Regular	Justified	n.	%
A	193	140	48	188	97%
E	69	31	28	59	86%
I	109	34	64	98	90%
O	100	51	26	77	77%
U	69	57	4	61	88%
	540	313	170	483	89%

TABLE 8: RESULTS COMBINED WITH CÁMARA-ARENAS 2018

We will now consider the teachability of these rules in the context of L2 pronunciation teaching. For that, we would encourage the exploration of those domain-specific rules such as V + C + <i…># and V+C+ vv…, which can be applied to a great number of PR3.2 irregular words with few exceptions themselves (see tables 4 and 5 in the Results section).

Medial <v> was applied to four words (5 considering *civic,* which we classified as having a V + C + <i…># context*)* in the domain of <i>, while this might seem insufficient we could argue that the rule has very few exceptions (6) against the 33 regular words in Cámara-Arenas (2010, 131-132). This rule works well with the fact that all <i> exceptions had the alternative version of the vowel, so there is no doubt of which vocalic sound to pronounce in an irregular PR3.2. word. When an irregular stressed vowel is followed by the medial consonant <v> there is a tendency for that vowel to be pronounced according to its alternative version. There are many possible reasons behind this, among them, the likely confusion that doubling the consonant, <vv>, might have caused in the early stages of printing. Considering the almost complete absence of double <v> (*chivvy* comes to mind as the only case) in the English spelling system, we would think that the few cases in which the preceding vowel is followed in its nominal version are, in fact, exceptions. So, instead of considering *ever* as an irregular PR3.2. word, we might have to consider *fever* as an irregular word that could belong to a different rule. The case of medial <v> leaves many possibilities yet for the development of domain-specific rules.

Medial <x> similarly represents a small number of our words but has been found to have very high reliability in the literature (Cámara-Arenas 2010, 99 for <e> and 131 for <i>) and would be easily remembered in a L2 teaching situation. <x> is often

considered as a "cc" cluster, since it always represents two sounds, so an alternative version of the preceding vowel is justified.

The False Proparoxytone rule is for words which visually may give the impression of having the stressed vowel in the second to last syllable. This would make them proparoxytones, hence the name. It is specific to the domains of <a> (Cámara-Arenas 2010, 65) and <e> (95). Although we have found some examples also in <i> and <o>. A domain-specific rule relies on the number of words which subscribe to it being larger than its exceptions. Some of these exceptions can be explained through other domain-specific rules, (Cámara-Arenas 2010, 100). However, for the case of <i> and <o> this rule is not advisable. Despite having *image* or *figure*, we have to consider the existence of *climate* or *silence.* Conversely, <o> has a similar diversity in the possible pronunciations of false proparoxytone words (163): *solace* but *opiate.*

The only pre-nuclear domain-specific rule we have found for our set of exceptions was initial <w> for the domain of <a>. This rule like all initial specification rules is dependent on post-nuclear rules, either general or domain-specific. We have /swam/ (*swam),* but /wɒsp/ (*wasp)* and /wɔːl/ (*wall)*, all of them PR2.1. irregular words.. And we have /ˈwɔːtə/ (*water,* a PR3.2. exception in our list) but /ˈweɪvə/ (*waver,* a PR3.2. regular word). Graphonemic indicators in pre-nuclear position are somewhat unreliable when considered in isolation.

All in all, we believe that we have proved the effectivity of domain-specific rules for the most problematic of all Pronunciation Rules as established by Cámara-Arenas (2018), with an increased efficiency of almost 89% for PR3.2. This was done by considering all domain-ruled words as regular. Taking into account a teaching perspective, we could only consider only the most reliable rules. Counting only those for which we have found justification in the literature (no exceptions and no tentative classifications), the regularity of PR3.2. would drop to 83% (see table 9 below). This percentage is still a lot higher than PR3.2.'s initial regularity of 58% and agrees with the 83.6% general predictability found in the system (Cámara-Arenas 2018, 207).

	Cámara-Arenas 2018		Justified in Cámara-Arenas 2010	
	PR3.2. Total	**Regular**	**n.**	**%**
A	193	140	42	94%
E	69	31	23	78%
I	109	34	54	81%
O	100	51	17	68%
U	69	57		83%
	540	313	136	83%

TABLE 9: REGULARITY OF PR3.2. WITH RULES JUSTIFIED IN THE LITERATURE

7. CONCLUSION

With a final number of 49 unjustified, purely irregular words from the initial 219, we can easily see that the perceived irregularity of Pronunciation Rule 3.2 is in fact not as pronounced. With the application of domain-specific rules we have been able to justify the irregular pronunciation of most exceptions to the general rule. Domain-specific rules need not be too complicated or numerous. We have proved that, in fact, with the application of only three simple rules (v+c+vv for *posi**t**ion*, v+c+<i…>*#* for *gen**e**tic* and the false proparoxytone rule for *m**a**nage*) more than half of the irregular PR3.2. words become regular. We need not delve too deep in casuistic or etymology to group these regularized words, as most of them exhibit a very clear common structure.

These alternative pronunciations to PR3.2. are not too complex either. Most unigraphs present always the short version (called alternative pronunciation) of the vowel instead of rule 3.2.'s prediction of the nominal version of the vowel. For the few special cases in which a vowel is pronounced neither with the nominal version nor the alternative short one, the word is usually well-known (e.g. *woman*, *sugar* or *any)*.

We have brought up the regularity of PR3.2. words to a degree that is level with all other pronunciation rules. With the application of domain-specific rules, the pronunciation of the stressed vocalic unigraph in paroxytone words with one medial consonant is 83% regular. This way, paroxytone words are comparable with all other structures in the system.

There are interesting possibilities for future domain-specific rules in some of the contexts that we have alluded to but not developed. Possibly, in the future, with a larger set of words we might be able to develop wider rules to encompass each domain's particular features. Furthermore, we would also like to polish the code written for the classification of words, improving its substitution rules so it might serve this same purpose with a larger database of rarer, more diverse words. With increased input the program might also work to find common patterns among the words, probably even selecting statistical features that we would overlook when doing it manually.

Some of the domain-specific rules we have expressed in this study would prove useful when explaining the irregularities found for the other sets of Pronunciation Rules. It is to be expected that in much the same way we had some elided vowels responsible for the misclassification of proparoxytones, some of our PR3.2. words might be unjustly classified as exceptions to other rules, quite probably for those belonging to the PR2 and PR3 groups. For this very same reason, we might also enjoy the different classification that using another dictionary would give us, since Carnegie has proved to be quick to embrace changes in pronunciation. A contrast between classifications made through different dictionaries might shed light on linguistic change and the acceptance of neologisms.

Lastly, we believe that with this study we have managed to create a preliminary structure and methodology that has taken a good step in the direction of developing automatic tools for grapho-phonemic studies. An accomplishment which might prove essential for the regularization and understanding of English' phoneme to grapheme mappings and their application to contexts of ELF pronunciation teaching.

8. REFERENCES

8.1. DIRECT REFERENCES:

Ausín, Adolfo, and Megan Sutton. "An L2 Pronunciation Judgment Task." *Selected Proceedings*, edited by Claudia Borgonovo, 2010, pp. 234–45.

Bakla, Arif, and Mehmet Demizeren. "An Overview of the Ins and Outs of L2 Pronunciation: A Clash of Methodologies." *Journal of Mother Tongue Education*, vol. 6, no. 2, 2018, pp. 475–95.

Bozman, Timothy. *Sound Barriers. A Practice Book for Spanish Students of English Phonetics.* Universidad de Zaragoza, 1988.

Burri, Michael. "Student Teachers' Cognition about L2 Pronunciation Instruction: A Case Study." *Australian Journal of Teacher Education*, vol. 40, no. 10, Oct. 2015, pp. 66–87.

Cámara-Arenas, Enrique. "EFL Grapho-Phonemics: The 'Teachability' of Stressed Vowel Pronunciation Rules." *Journal of the Spanish Association of Anglo-American Studies*, vol. 40.2, Dec. 2018, pp. 197–218.

---. "'Graphonemic Indicators' of Vowel Pronunciation: Suggestions Fot Research and Teaching." *Estudios de Metodología de La Lengua Inglesa.*, edited by Leonor Pérez-Ruiz et al., vol. 4, 2008, pp. 91–100.

---. *La Vocal Inglesa*. *Correspondencias Grafo-Fonémicas.* Secretariado de Publicaciones. Universidad de Valladolid, 2010.

Castles, Anne, et al. “Variations in Spelling Style among Lexical and Sublexical Readers.” *Journal of Experimental Child Psychology*, vol. 64, 1997, pp. 98–118.

Cummings, D. W. *American English Spelling*. Johns Hopkins University Press, 1988.

Derwing, Tracey M. *Putting an Accent on the Positive: New Directions for L2 Pronunciation Research and Instruction*. 2018, pp. 12–18.

Deterding, David, and Christine Lewis. “Pronunciation in English as Lingua Franca.” *Second Handbook of English Language Teaching*, Springer, 2019, pp. 1–15.

Eddington, David, et al. “Syllabification of American English: Evidence from a Large-Scale Experiment. Part I.” *Journal of Quantitative Linguistics*, vol. 20, no. 1, 2013, pp. 45–67.

---. “Syllabification of American English: Evidence from a Large-Scale Experiment. Part II.” *Journal of Quantitative Linguistics*, vol. 20, no. 2, 2013, pp. 75–93.

Escudero-Mancebo, David, et al. *Analysis of the Efficiency of Repeating Activities for Improving Prosody in L2 Pronunciation Training*. 2018, pp. 299–303.

Fry, Edward. “Phonics: ALarge Phoneme-Grapheme Frequency Count Revised.” *Journal of Literacy Research*, vol. 36, no. 1, 2004, pp. 85–98.

Hayes, Heather, Rebecca Treiman, et al. “Children Use Vowels to Help Them Spell Consonants.” *Journal of Experimental Child Psychology*, vol. 94, 2006, pp. 27–42.

Hayes, Heather, Brett Kessler, et al. “English Spelling: Making Sense of a Seemingly Chaotic Writing System.” *Perspectives, The International Dyslexia Association*, Summer 2015.

Isaacs, Talia. “Integrating Form and Meaning in L2 Pronunciation Instruction.” *TESL Canada Journal*, vol. 27, no. 1, Winter 2009, pp. 1–12.

Jenkins, Jennifer. “Pedagogic Priorities 2: Negotiating Intelligibility in the ELT Classroom.” *The Phonology of English as an International Language*, Oxford University Press, 2000, pp. 63–93.

---. “Proposals for Pronunciation Teaching for EIL.” *The Phonology of English as an International Language*, Oxford University Press, 2000, pp. 202–39.

Jenkins, Jennifer, and Constant Leung. “Assessing English as a Lingua Franca.” *Language Testing and Assessment, Encyclopedia of Language and Education*, Springer International Publishing, 2017, pp. 1–16.

Kessler, Brett, and Rebecca Treiman. “Is English Spelling Chaotic? Misconceptions Concerning Its Irregularity.” *Reading Psychology*, vol. 24, no. 3–4, Dec. 2003, pp. 267–89.

Lass, Roger. *English Phonology and Phonological Theory*. Cambridge University Press, 1976.

Levis, John, and Alene Moyer. *Future Directions in the Research and Teaching of L2 Pronunciation*. 2014, pp. 275–91.

Nguyen, Anh Duc Dao. “Learner Perception of L2 Pronunciation Instruction.” *The English Teacher*, vol. 47, no. 2, 2018, pp. 44–52.

O’Neal, George. “The Accomodation of Intelligible Segmental Pronunciation. Segmental Repairs and Adjustments in English as a Lingua Franca Interactions.” *Journal of Second Language Pronunciation*, vol. 5, no. 1, 2019, pp. 119–38.

Pacton, Sébastien, et al. "Children Benefit from Morphological Relatedness Independently of Orthographic Relatedness When They Learn to Spell New Words." *Journal of Experimental Child Psychology*, vol. 171, Mar. 2018, pp. 71–83.

Procter, P. *Longman Dictionary of Contemporary English*. 1978th ed., Longman.

Robbins, Kelly P., et al. "Assessing Specific Grapho-Phonemic Skills in Elementary Students." *Assessment for Effective Intervention*, vol. 36, no. 1, 2010, pp. 21–34.

Saito, Kazuya. "Corrective Feedback and the Development of L2 Pronunciation." *The Cambridge Handbook of Corrective Feedback in Language Learning and Teaching.*, edited by H. Nassaji and E. Kartchava, Cambridge University Press, in press.

Treiman, Rebecca, et al. "Context Sensitivity in the Spelling of English Vowels." *Journal of Memory and Language*, vol. 47, 2002, pp. 448–68.

Treiman, Rebecca, and Kelly Boland. "Graphotactics and Spelling: Evidence from Consonant Doubling." *Journal of Memory and Language*, vol. 92, 2017, pp. 254–64.

Treiman, Rebecca, and Sloane Wolter. "Phonological and Graphotactic Influences on Spellers' Decisions about Consonant Doubling." *Memory & Cognition*, vol. 46, no. 4, May 2018, pp. 614–24.

Venezky, Richard L. *The American Way of Spelling*. The Guildford Press, 1999.

Wijk, Axel. *Rules of Pronunciation for the English Language*. Oxford University Press, 1966.

Zoghbor, Wafa. “Revisiting English as a Foreign Language (EFL) vs English Lingua Franca (ELF): The Case for Pronunciation.” *Intellectual Discourse*, vol. 26, International Islamic University Malaysia, 2018, pp. 829–58.

8.2. INDIRECT REFERENCES:

DeKeyser, R. M. “Beyond Focus on Form: Cognitive Perspectives on Learning and Practical Second Language Grammar.” *Focus on Form in Classroom Second Language Acquisition*, Cambridge University Press, 1998, pp. 42–63.

Fudge, E. C. “Syllables.” *Journal of Linguistics*, vol. 5, 1969, pp. 253–86.

APPENDIX

PYTHON CODE

PREPARATION: GENERATE_WORDTRANS

```
"""
Script to create a joint csv file with the words to be
analysed and the phonetic translations.

How to run:
  >>> python generate_wordtrans.py input_words
input_phonetic output_filename

Required:
- input_words: file with a list of words, each on a new
line
- input_phonetic: file with the phonetic translations of a
series of words. Formatted like this:
        ABACUS  AE1 B AH0 K AH0 S
        word + 2 spaces + 1st sound + 1 space + 2nd sound
+ space + ...
        The stressed sound is marked with a 1 after the
sound.
        A non stressed vocalic sound is marked with a 0
after the sound.

It will generate:
- output_filename
        Each line corresponds to a different word.
        word; ['sound1', 'sound2', ..., 'soundn'];
[stressed1, stressed2, ..., stressedn]
        where stressed can be:
            2: vocalic sound + secondary
            1: vocalic sound + primary
            0: vocalic sound + not stressed
            -1: not vocalic sound


Example:
  >>> python generate_wordtrans.py "words.csv"
"phontransCAR" "mydict.csv"
  or, for default values:
  >>> python generate_wordtrans.py


"""
```

```
import sys        # to access the cmd line arguments
import csv


class Words():

    def __init__(self):

        self.words = []
        self.phon = []
        self.stressed = []

        return

class PhonTrans():

    def __init__(self):

        self.words = []
        self.phon = []
        self.stressed = []

        return


def load_words(filename, wordsclass):
    # Open file
    i = 0
    with open(filename, encoding='utf-8') as csv_file:
        csv_reader = csv.reader(csv_file, delimiter=',',
dialect=csv.excel)
        for row in csv_reader:

wordsclass.words.append((row[0]).lower().strip())
            wordsclass.phon.append("")
            wordsclass.stressed.append("")
            i += 1

    return wordsclass

def load_phon(filename, phonclass):
    #open file with carnegie trancriptions
    i = 0
    with open(filename, encoding='utf-8') as csv_file:
        csv_reader = csv.reader(csv_file,
dialect=csv.excel)
        for row in csv_reader:
```

```
            aux = row[0].split('  ')
            phonclass.words.append(aux[0].lower())
            # Separate the information of the translation
on phon and stressed
            aux_phon = []
            aux_stressed = []
            sounds = aux[1].split(' ')
            for j in range(len(sounds)):    # for each
sound

                if sounds[j][-1] == "1":
                    # This sound is stressed
                    aux_stressed.append(1)
                    aux_phon.append(sounds[j][:-1])     #
save everything except the 1 at the end
                elif sounds[j][-1] == "2":
                    #this is the secondary stress
                    aux_stressed.append(2)
                    aux_phon.append(sounds[j][:-1])
                elif sounds[j][-1] == "0":
                    # This sound is not stressed
                    aux_stressed.append(0)
                    aux_phon.append(sounds[j][:-1])     #
save everything except the 0 at the end
                else:
                    # not vocalic sounds
                    aux_stressed.append(-1)
                    aux_phon.append(sounds[j])          #
save everything

            phonclass.phon.append(aux_phon)
            phonclass.stressed.append(aux_stressed)
            i += 1

    return phonclass

def merge_words_phon(wordsclass, phonclass):
    errors_list= []
    for i in range(len(wordsclass.words)):
        # For each file from the words file...
        # look if there is a translation in the phon file
        try:
            j = phonclass.words.index(wordsclass.words[i])
        except ValueError as _e:
            # in case there is no corresponding word on
the phonetic dictionary
            errors_list.append(i)
```

```
            print ("    * Word not found in dictionary: ",
wordsclass.words[i])
            continue    # pass, go check the next word

        wordsclass.phon[i] = phonclass.phon[j]
        wordsclass.stressed[i] = phonclass.stressed[j]
    #delete errors from wordsclass
    for e in errors_list[::-1]:
        wordsclass.words.pop(e)
        wordsclass.phon.pop(e)
        wordsclass.stressed.pop(e)


    return wordsclass

def save_csv_wordsplus(wordsclass, filename):

    with open(filename, mode='w') as outfile:
        csv_writer = csv.writer(outfile, delimiter=';',
lineterminator='\n')

        for i in range(len(wordsclass.words)):
            csv_writer.writerow([wordsclass.words[i],
wordsclass.phon[i], wordsclass.stressed[i]])

    return


if __name__ == '__main__':

    # Get arguments from function call

    # 1st argument: name of the script
    # 2nd argument: input_words
    # 3rd argument: input_phonetic
    # 4th argument: output_filename

    if len(sys.argv) == 4:
        words_file = sys.argv[1]
        phon_file = sys.argv[2]
        out_file = sys.argv[3]
    else:
        # Try using default names
        print(">> No specific input or output files given.
File will be run with default names...")
        words_file = "words.csv"
        phon_file = "phontransCAR"
        out_file = "mydict.csv"
```

```
    # Initialize classes
    words = Words()
    phon = PhonTrans()

    print(">> Loading words in
'{0}'...".format(words_file))
    words = load_words(words_file, words)

    print(">> Loading phonetic translations in
'{0}''...".format(phon_file))
    phon = load_phon(phon_file, phon)

    print(">> Merging words and phonetic translations...")
    words = merge_words_phon(words, phon)

    print(">> Saving words and corresponding phonetic
translations to '{0}'...".format(out_file))
    save_csv_wordsplus(words, out_file)


    print(">> End of program")
```

SELECTION: MAIN_NEW_SIMPLE

```
vowels = 'aiou'
exceptions = 'lrwyhe'
regular = ['EY', 'IY', 'AY', 'OW', 'UW', 'YUW']
#Exceptions might need to be changed. Different
positions/combinations.

import csv


class MyDict():

    def __init__(self):

        self.words = []
        self.phon = []
        self.stressed = []
        self.idx = []
        self.valid = []
        self.form = []
        self.paroxytones = []

        return

    def load_words(self, filename):
```

```
        # Open file
        i = 0
        with open(filename, encoding='utf-8') as csv_file:
            csv_reader = csv.reader(csv_file,
delimiter=';', dialect=csv.excel)
            for row in csv_reader:
                self.words.append(row[0])

                aux = row[1][1:-1].split(',')
                for j in range(len(aux)):
                    aux[j] = aux[j].strip('
').replace("'", "")
                self.phon.append(aux)

                aux = row[2][1:-1].split(',')
                for j in range(len(aux)):
                    aux[j] = int(aux[j])
                self.stressed.append(aux)

                self.idx.append(i)
                self.valid.append(False)
                self.form.append('')
                self.paroxytones.append(False)
                i += 1
        return

    def formulae(self):
        """ Calculates the grapho-phonemic formula of the
given word. """

        for i in self.idx:
            form = "$"
            word = self.words[i]

            # For each word...
            # Change ch for c
            if 'ch' in word:
                idx = word.find('ch')
                word = word[:idx] + 'c' + word[idx+2:]
            #GU-as a consonant
            if 'gu' in word:
                idx = word.find('gu')
                if word[idx+2] in vowels:
                    word = word[:idx] + 'c' + word[idx+2:]

            # Check if y/w at the end, change for vowel
            if word[-1] == 'y':
```

```
                word = word[:-1] + 'a'      # any vowel
works
            if word[-1] == 'w':
                word = word[:-1] + 'a'      # any vowel
works

            # exceptions in non-exceptional positions:
initial
            if word[0] == 'r':
                word = 'c' + word[1:]
            if word[0] == 'l':
                word = 'c' + word[1:]
            if word[0] == 'w':
                word = 'c' + word[1:]

            # v for vowel, c for consonant, special
letters
            for letter in word:
                if letter in vowels:
                    form += 'v'
                elif letter in exceptions:
                    form += letter
                else:
                    form += 'c'


            form += '#'

            # 2 vowels together = diphthong
            for j in range(len(form))[::-1]:
                if (form[j-1:j+1] == 'vv'):
                    form = form[:j-1] + 'd' + form[j+1:]
                if(form[j-1:j+1] == 'ev') or (form[j-
1:j+1] == 've'):
                    form = form[:j-1] + 'd' + form[j+1:]
                if (form[j-1:j+1] == 'ee'):
                    form = form[:j-1] + 'd' + form[j+1:]

            # Save form to class
            self.form[i] = form

        return

    def validate(self):
        """ self.valid is True if the given word is two
syllabic or more. """

        for i in self.idx:
```

```
            out = False
            form = self.form[i]
            if form.count('v') + form.count('d') +
form.count('e') >= 2:
                out = True

            self.valid[i] = out
        return

    def print_monosyllabic(self):
        """ Print 4 columns (index, word, form, count) for
only the monosyllabic words.
        """
        # Print only mono-syllabic words
        print("\n{0:<10} {1:<10} {2:<20}
{3:<10}".format("idx", "word", "formula", "count"))
        print("-----------------------------------------
-------------------------------")
        mono_count = 0
        for i in self.idx:
            if self.valid[i] == False:
                mono_count += 1
                print("{0:<10} {1:<10} {2:<20}
{3:<10}".format(self.idx[i], self.words[i], self.form[i],
mono_count))

        return

    def print_valid(self):
        """ Print 4 columns (index, word, form, count) for
only the valid words.
        More than 1 syllable.
        """
        # Print only valid words
        print("\n{0:<10} {1:<20} {2:<30}
{3:<10}".format("idx", "word", "formula", "count"))
        print("-----------------------------------------
-------------------------------")
        mono_count = 0
        for i in self.idx:
            if self.valid[i] == True:
                mono_count += 1
                print("{0:<10} {1:<20} {2:<30}
{3:<10}".format(self.idx[i], self.words[i], self.form[i],
mono_count))

        return
```

```
    def print_paroxytones(self):
        """ Print 5 columns (index, word, form, stress,
count) for only the paroxytone words.
        Stress in penultimate syllable.
        """
        # Print only paroxytone words
        print("\n{0:<10} {1:<10} {2:<20} {3:<30}
{4}".format("idx", "count", "word", "formula",
"stressed"))
        print("-----------------------------------------
-------------------------------------------------
")
        mono_count = 0
        for i in self.idx:
            if self.paroxytones[i] == True:
                mono_count += 1
                print("{0:<10} {1:<10} {2:<20} {3:<30}
{4}".format(self.idx[i], mono_count, self.words[i],
self.form[i], self.stressed[i]))
        return

    def find_clusters(self, clusters, valid=False,
paroxytones=False):
        """
        Valid=True
        To print only valid words in the cluster (PR3.2
words)

        Paroxytones=True
        To print only paroxytones containing the cluster:
valid words cannot be followed by another consonant.


        """

        idx_cluster = []

        for i in self.idx:

            if (self.valid[i] == False) & (valid==True):
                continue
            if (self.paroxytones[i] == False) &
(paroxytones==True):
                continue

            # Here only if word is valid
            for cluster in clusters:
```

```
                if cluster in self.form[i]:
                    idx_cluster.append(i)

        print("    {0} words
found.".format(len(idx_cluster)))
        return idx_cluster

    def find_paroxytone(self):
        """ self.paroxytone is True if the given word has
its stressed vowel in the penultimate syllable. """
        for i in self.idx:
            out = False
            possible = False
            aux = self.stressed[i]
            for j in range(len(aux))[::-1]:
                if aux[j]== -1:
                    continue
                elif possible== True:
                    #previous sound unstressed
                    if aux[j]== 1:
                        out = True
                        break
                    else:
                        possible = False
                        break
                elif aux[j] != 1:
                    possible = True
                else:
                    #last vocalic sound is stressed:>
oxytonic
                    break
                #here only if last vocalic sound is
unstressed (requirement)
            self.paroxytones[i] = out
        print("    {0} words
found.".format(self.paroxytones.count(True)))
        return

    def find_prediction(self):

        """
        To find the expected pronunciation in valid
paroxytones:
        EY1, IY1, AY1, OW1, UW1, YUW1
        """
        idx_prediction = []
        idx_exception = []
        for i in self.idx:
```

```
            if (self.valid[i] == False) or
(self.paroxytones==False):
                continue
            # Here only if word is valid and paroxytone
            j = self.stressed[i].index(1)
            sound = self.phon[i][j]
            if sound in regular:
                idx_prediction.append(i)
            else:
                idx_exception.append(i)
        print("    {0} regular words
found.".format(len(idx_prediction)))
        print("    {0} irregular words
found.".format(len(idx_exception)))
        return idx_prediction, idx_exception

    def print_prediction(self, idx_pred):
        """ Print 5 columns (count, index, primary, word,
form) for only the words with index in the given vector.
        """
        print("\n{0:<10} {1:<10} {2:<10} {3:<25}
{4:20}".format("count", "idx", "primary", "word",
"formula"))
        print("-----------------------------------------
-----------------------------------------------------
-----------------------------------------------------
---------------------------------")
        word_count = 0
        for i in self.idx:

            if i in idx_pred:
                word_count += 1
                # Generate strings from predictable valid
paroxytone words
                j = self.stressed[i].index(1)
                primary = self.phon[i][j]
                print("{0:<10} {1:<10} {2:<10} {3:<25}
{4:20}".format(word_count, self.idx[i], primary,
self.words[i], self.form[i]))

        return

    def print_selection(self, idx_sel):
        """ Print 6 columns (count, index, word, form,
stress, phontrans) for only the words with index in the
given vector.
        """
```

```
        print("\n{0:<10}; {1:<10}; {2:<20}; {3:<35};
{4:<50}; {5}".format("count", "idx", "word", "formula",
"stress", "phontrans"))
        print("-----------------------------------------
-----------------------------------------------------
-----------------------------------------------------
---------------------------------")
        word_count = 0
        for i in self.idx:

            if i in idx_sel:
                word_count += 1
                # Generate strings from the stressed and
phon lists
                aux_stressed = ""
                aux_phon = ""
                for j in range(len(self.stressed[i])):
                    aux_stressed = aux_stressed +
str(self.stressed[i][j])+ ", "
                    aux_phon = aux_phon + self.phon[i][j]
+ ' '
                print("{0:<10}; {1:<10}; {2:<20}; {3:<35};
{4:50}; {5:80}".format(word_count, self.idx[i],
self.words[i], self.form[i], aux_stressed, aux_phon))

        return

    def get_indexes(self, idx1, idx2, fun='and'):

        """
        Two vectors, will cross-reference both (clusters
and expected pronunciation).
        Returning an index
        """

        idx= []

        for i in idx1:
            if i in idx2:
                #index is in both vectors
                idx.append(i)

        return idx

if __name__ == '__main__':


    filename = 'mydict.csv'
```

```
    # Open file mydict
    print('>> Load words from file...')
    mydict = MyDict()
    mydict.load_words(filename)

    # Generate formulae from words
    print('>> Calculate formulae from words...')
    mydict.formulae()

    # Calculate valid words
    print('>> Validate 2 syllabic words...')
    mydict.validate()
    #mydict.print_valid()
    #mydict.print_monosyllabic()

    # Find different clusters


    #Find stress pattern
    print('>> Finding paroxytones...')
    mydict.find_paroxytone()
    #mydict.print_paroxytones()

    print('>> Finding different clusters...')
    #these are the three parts of PR 3.2.
    idx_cluster_a = mydict.find_clusters(clusters=['vcv',
'ece', 'ecv', 'vce', 'vhv', 'ehv', 'ehe', 'vhe', 'vlv',
'elv', 'ele', 'vle', 'vrv', 'erv', 'ere', 'vre', 'vcd',
'ecd', 'vhd', 'ehd', 'vld', 'eld', 'vrd', 'erd' ],
valid=True, paroxytones=True)
    #mydict.print_selection(idx_cluster_a)

    #idx_cluster_b =
mydict.find_clusters(clusters=['vcrv', 'ecrv', 'vcre',
'ecre'], valid=True, paroxytones=True)
    #mydict.print_selection(idx_cluster_b)

    #idx_cluster_c =
mydict.find_clusters(clusters=['vclv', 'eclv', 'vcle',
'ecle'], valid=True, paroxytones=True)
    #mydict.print_selection(idx_cluster_c)


    #Find expected pronunciation in all paroxytones
    #print('>> Finding the expected pronunciation...')
    [idx_prediction, idx_exception] =
mydict.find_prediction()
```

```
    #mydict.print_prediction(idx_prediction)


    #PR3.2.a
    #print('>> Finding the expected pronunciations of
PR3.2.a ...')
    #regularwords=mydict.get_indexes(idx_cluster_a,
idx_prediction)
    #print("    {0} words
found.".format(len(regularwords)))
    #mydict.print_selection(regularwords)
    print('>>Finding exceptions to PR3.2.a ...')
    irregularwords=mydict.get_indexes(idx_cluster_a,
idx_exception)
    print("    {0} words
found.".format(len(irregularwords)))
    mydict.print_selection(irregularwords)

    #PR3.2.b
    #print('>> Finding the expected pronunciations of
PR3.2.b ...')
    #regularwords=mydict.get_indexes(idx_cluster_b,
idx_prediction)
    #print("    {0} words
found.".format(len(regularwords)))
    #mydict.print_selection(regularwords)
    #print('>>Finding exceptions to PR3.2.b ...')
    #irregularwords=mydict.get_indexes(idx_cluster_b,
idx_exception)
    #print("    {0} words
found.".format(len(irregularwords)))
    #mydict.print_selection(irregularwords)

    #PR3.2.c
    #print('>> Finding the expected pronunciations of
PR3.2.c ...')
    #regularwords=mydict.get_indexes(idx_cluster_c,
idx_prediction)
    #print("    {0} words
found.".format(len(regularwords)))
    #mydict.print_selection(regularwords)
    #print('>>Finding exceptions to PR3.2.c ...')
    #irregularwords=mydict.get_indexes(idx_cluster_c,
idx_exception)
    #print("    {0} words
found.".format(len(irregularwords)))
    #mydict.print_selection(irregularwords)
```