Minha Terra Tem _____: Patterns of Text Reuse in "Song of Exile" and its Intertexts

Joshua Alma Enslen

WEST POINT | UNITED STATES MILITARY ACADEMY
ORCID: 0009-0002-3858-0218

Jocelyn R. Bell

HOBART AND WILLIAM SMITH COLLEGES
ORCID: 0000-0002-3623-7865

RESUMO

Um texto fundamental do Romantismo Brasileiro, a "Canção do Exílio" de Antônio Gonçalves Dias, escrita em Coimbra (1843) e publicada em Niterói (1846), se posiciona de forma única como o poema mais imitado do Brasil. Com base em um conjunto de dados de 500 intertextos abrangendo os anos 1846 a 2015, este estudo contempla a natureza do empréstimo textual encontrado nas variações do poema. Indo além das palavras compartilhadas entre os textos para os padrões sintáticos em jogo, este trabalho define e discute de forma detalhada o conceito de "Modelos Sintáticos", desenvolve um algoritmo para a sua identificação e considera a sua utilidade para compreender este problema específico de reuso de texto.

PALAVRAS-CHAVE

Antônio Gonçalves Dias; Romantismo brasileiro; reuso de texto; intertextualidade; modelos sintáticos

ABSTRACT

A foundational text of Brazilian Romanticism, Antônio Gonçalves Dias's "Canção do Exílio," written in Coimbra (1843) and published in Niterói (1846), is uniquely positioned as Brazil's most imitated poem. Based on a data set of 500 intertexts spanning the years 1846-2015, this study contemplates the nature of textual borrowing found in the poem's variations. Looking beyond words shared across texts to the syntactic patterns at play, this paper defines and discusses in detail the concept of Syntactic Templates, develops an algorithm for their identification and considers their utility for understanding this specific text reuse problem.

KEYWORDS

Antônio Gonçalves Dias; Brazilian Romanticism; text reuse; intertextuality; syntactic templates

A foundational text of Brazilian Romanticism, Antônio Gonçalves Dias's "Canção do Exílio," written in Coimbra (1843) and published in Niterói (1846), is uniquely positioned as Brazil's most imitated poem.[1] Based on a data set of 500 intertexts spanning the years 1846-2015, this study contemplates the nature of textual borrowing found in the poem's variations. Looking beyond words shared across texts to the syntactic patterns at play, this paper defines and discusses in detail the concept of Syntactic Templates, develops an algorithm for their identification and considers their utility for understanding this specific text reuse problem.

In his recent book, *Song of Exile: A Cultural History of Brazil's Most Popular Poem, 1846-2018*, Joshua Enslen introduces the concept of Syntactic Templates in the following manner:

> Syntactic Templates are grammatical structures established in ["Canção do Exílio"], such as "Minha terra tem palmeiras," from which variations typically generate related texts. In these templates, the syntax of the original remains intact while Significant Words are altered, such as [. . .] "Minha Dinda tem cascatas,"[2] as "Dinda" and "cascatas" replace "terra" and "palmeiras" (Enslen, 2022: 6)

Exploring the cultural relevance of what Manuel Portela described in his afterword as "the most remixed Brazilian poem" (Portela, 2022: 157), Enslen's book demonstrates how Syntactic Templates function as primary generators of variation within the corpus of 500 intertexts. The substantive chapters of Enslen's study compile for analysis lists of terms based on template variation from corpora derived from specific periods in Brazilian history, such as the military dictatorship (1964-1985) or the First Republic through the Vargas regime (1889-1945). These terms replace Significant Words in the original and, in doing so, introduce new themes and ideas into the corpus lexicon.[3] The book uses these lists as a point of departure for

---

1 Cf. Gonçalves Dias, 1846: 9-10

2 Cf. Soares,1992: 15

3 "*Significant Words*, based on word frequency analysis, are the most repeated nouns among the intertexts, which together outline the general structure of a network. These *Significant Words* establish nodes of contact with the original and across the intertexts wherever they appear. The three most repeated nouns among all the texts are 'terra,' 'palmeiras,' and 'sabiá.'" (Enslen, 2022: 5-6)

"identify[ing] the contours of a vast network of influence that carries at its core questions of Brazilian history, identity, and culture" (Enslen, 2022: 14).

Although this first attempt to define Syntactic Templates represents a reasonable argument for their usefulness, the book's discussion is technically insufficient since the conceptual nature of these grammatical structures is not specified in exact terms, nor their complexity fully explored. Enslen also argues, "In effect, each word in the poem represents a specific link in a chain of words that, when strung together, constitute the original. In the poem's intertexts, portions of this sequence of words remain identifiable in varying degrees of modification" (Enslen, 2022: 5). Further refining the book's focus on intertextuality as a preoccupation with words "as minimal textual unit" (Kristeva, 1986: 37) and as strings of symbols, this paper examines more closely the underlying, if implied, instructions used in generating these strings of symbols, or Syntactic Templates. This paper's view of "Canção do Exílio" shifts the focus from words to algorithms, both abstract and concrete.

Focusing on the first line of the poem, "Minha terra tem palmeiras," this paper will expand on Enslen's definition of the term Syntactic Template, clarifying the concept to facilitate (through the use of an algorithm) the analysis of even greater quantities of intertexts and with a more pronounced analytical rigor. This deep investigation of the first line is justified both thematically and statistically. In Chapter 2, Enslen offers insight into the prominence of first-line variations among the intertexts, describing this oft repeated and varied line in post-colonial terms:

> The first three words of 'Canção do exílio,' 'Minha terra tem,' grounded in the discourses of colonization, exploration, and the acquisition of land and power, configure prominently in the dialectics of Brazil's post-colonial identity. . . The phrase's syntactical structure. . . establishes the role of all nouns in the original as signifiers for what the land or, by extension, the nation is. (Enslen, 2022: 15-16).

Likewise, a numerical justification for this paper's focus on the first line of the poem can be surmised with the following observations. The first line contains two of the three most repeated nouns in the 500 texts: "terra," "sabiá," and "palmeiras" appear 1245, 417 and 372 times respectively (Enslen, 2022: 204). In the original poem, G. Dias repeats the phrase verbatim two times in lines 11 and 17. Then, its syntax is closely imitated in lines 5 thru 8 and it is repeated again in line 13 with the word "primores" replacing "palmeiras." The first line and its variations (e.g., "nosso céu tem mais estrelas") constitute one-third of the 24 lines in the original poem. Additionally, the line contains the most frequently repeated 4-gram and 3-gram among the 500 texts in question. "Minha terra tem palmeiras" repeats a

total of 155 times in 106 texts and "minha terra tem" repeats verbatim a total of 615 times in 307 texts. Prompting the title of this paper, this 3-gram is a *de facto* formula for generating thematic variation among the intertexts since, as will be discussed, the nouns and noun phrases that follow it (like in the word game "Mad-Libs") complete the puzzle for what constitutes Brazilian identity: "Minha terra tem _____."

The immense intertextuality found within the corpus and its general reliance on the original poem approximate a text reuse problem. John Lee, in his 2007 article, "A Computational Model of Text Reuse in Ancient Literary Texts," provides a straightforward definition of the term: "Text reuse is the transformation of a source text into a target text in order to serve a different purpose" (Lee, 2007: 472). Building upon Lee's definition, Daria Elagina's recent article on Ethiopian and Eritrean literary traditions, describes the many forms text reuse takes: "Text reuse, which consists in the implicit or explicit repetition of text, is attested in a variety of forms, including quotations, allusions, paraphrases, and cross-linguistic text reuse" (Elagina, 2022: 1). All of these techniques of borrowing, whether manifested as parody, pastiche or some other genre, find a place in the Song of Exile corpus.

Uncovering instances of text reuse can be accomplished through close reading, but distant reading can lead to insights that otherwise might have been overlooked, especially with a large corpus. Several techniques and tools have been developed to establish text reuse as a distant reading practice. Perhaps the simplest tool for algorithmically detecting text reuse is the n-gram: a contiguous subsequence of n many words from a phrase or text, also known as a shingle of size n (Broder, 1997). The presence of common n-grams between texts can be used as evidence of intertextuality, including intertextuality between texts written in verse, as in Battles (2019). A related concept, termed an n-skip-gram, allows for "skips" between words in an n-gram. An n-skip-gram has an associated parameter called the maximum skip value, which is the largest number of consecutive words of a phrase that can be skipped over or ignored (Guthrie et al, 2006). For example, "minha tem palmeiras" is a 3-skip gram based on the phrase "minha terra tem palmeiras." If the maximum skip value is 0, then an n-skip-gram is the same as an n-gram. At times, texts are preprocessed before extracting n-grams and n-skip-grams; stop words are often filtered out and, in some cases, words are first reduced to their base form, as in Ganascia et. al (2014).

While n-grams are widely used as effective tools for comparison, at times the patterns of interest lie within the grammatical structure of phrases, rather than in the words themselves, a particularly relevant concept to this paper. A part-of-speech n-gram, or POS-n-gram, is formed similarly to an n-gram, except that in the place of words are parts of speech (Lioma & Rijs-

bergen, 2008). For example, the phrase "minha terra tem palmeiras" has the part-of-speech structure

$$S = [\text{Determiner, Noun, Verb, Noun}]$$

since "minha" is a determiner, "terra" and "palmeiras" are nouns, and "tem" is a conjugated form of the verb "ter." Any other contiguous subsequence of words in a text that fits this pattern would be counted as a POS-n-gram for this structure.

Adding another layer of complexity, in what are known as POS-n-skip-grams, as in n-skip-grams, there is a parameter that allows for "skips" or "jumps" over words up to the maximum skip value. This idea has been used to predict authorship of texts (Pokou et al, 2016). From "Canção do Exílio," an example of a POS-4-skip gram of the structure S above (with maximum skip value of 1) is "nosso céu tem estrelas." Of course, in "Canção do Exílio" the line is "nosso céu tem mais estrelas," but in the POS-4-skip gram the word "mais" has been ignored. This will become a crucial distinction in this paper.

A working definition of Syntactic Template, or simply a template, is an ordered list whose entries are sets of words or sets of sequences of words. Some of these sets will be created by grammatical categories, while others are more specific, at times containing only a single word. There is an important distinction between a template, which is a set of instructions for generating text, and the text generated by a template, or an instance of it. For a given template (an ordered list whose entries are sets of sequences of words), an instance of that template is a sequence of sequences of words satisfying that the first sequence of words in the instance is an element of the first entry of the template, the second sequence in the instance an element of the second entry of the template, and so on. By concatenating the smaller sequences that comprise an instance, the instance itself can be viewed as one sequence of words, and it is instances, not templates, that appear in texts. This paper will construct a hierarchy of templates, each more general than the last, with the phrase "Minha terra tem palmeiras" as an instance of each, and explore their adoptions in the data set of 500 intertexts as they trace themes relevant to Brazilian identity.

Many syntactic templates have "Minha terra tem palmeiras" as an instance. The relationships among the templates in the hierarchy are represented visually in Figure 1. In this figure, the represented phrases are instances found in at least one of the 500 texts. For example, the phrase "nossas vidas não têm mais amores," from Leandro Arruda's text bemoaning the squandering of Brazil's natural resources, is an instance of what this paper defines as a T_4 template (Arruda, 2012). The collections of instances

generated by the templates T_0 through T_6 are nested, in the sense that every instance of the T_1 template is also an instance of the T_2 template, every instance of the T_2 template is also an instance of the T_3 template, and so on. Yet, the converse is not always true: there are instances of the T_3 template that are not instances of the T_2 template, etc. The following line from Paulo Briguet's "Canção do Exíguo" is one such example: "minha terra não tem fé" (Briguet, 2009). All the phrases appearing in Figure 1 are cases of a T_i instance that is not a T_{i-1} instance (for i between 1 and 6).[4]
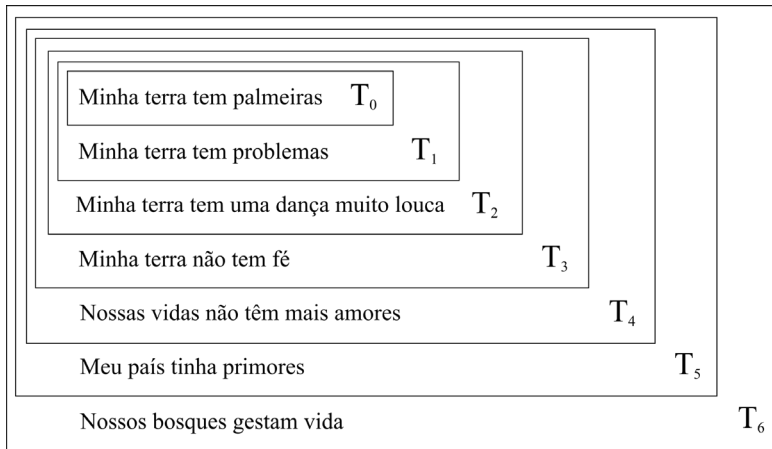

Figure 1

The first line of "Canção do Exílio," "Minha terra tem palmeiras," contains four words. The simplest syntactic template with four entries generating this phrase is the template with no moving parts, represented conceptually as

$$T_0 = [\{minha\}, \{terra\}, \{tem\}, \{palmeiras\}]$$

In this T_0 template, an ordered list with four slots, each entry is a set containing exactly one element, and the only possible instance of this template

---

4   The phrase T_1 appears in 5 separate texts among the 500, a telling fact since not just these five but many texts present the "problemas" found in Brazil. The phrase T_2 from Breno Lobato and Samara Cristina is found in an untitled text published on a blog in 2013, which playfully begins, "Minha terra tem o brega / Que faz o povo dançar / Super pop, o Águia de fogo / Faz a galera endoidar." Phrases T_3 and T_4 are discussed above. T_5 is from Miriam Panighel Carvalho's 2008 text entitled "Paráfrase da Canção do Exílio, de Gonçalves Dias," which was published online at *Recanto das Letras*. Phrase T_6 is from Luiz Alexandre Cruz Ferreira's text entitled "Ode ao Exílio," published in 2015 in an online newspaper from Santa Catarina, "O Alvoradense" (later taken down). Complete bibliographic information on texts represented in the figure can be found in Works Cited.

is "minha terra tem palmeiras." The 4-gram "minha terra tem palmeiras" appears in 21% of the 500 texts, which is evidence that n-grams (which are readily seen to be identical to instances of templates all of whose entries are one-word sets) are a powerful tool for expressing relatedness to the original. In what follows, syntactic templates will codify patterns beyond n-grams, to illuminate differences, in addition to similarities, expressed in the intertexts.

Compare the first line "Minha terra tem palmeiras" to the line "Minha terra tem empresas" from "Minha Terra Natal," an anonymous poem from 1853 about political and social problems in Bahia (Anonymous, 1853: 3). G. Dias's "palmeiras" have become "empresas," a key thematic difference, while still displaying significant textual overlap. This phenomenon occurs within the original itself, in the line "minha terra tem primores." A simple yet flexible template generating both of these phrases is an ordered list with four slots

$$T_1 = [\{minha\}, \{terra\}, \{tem\}, Noun]$$

where the first, second and third entries are the one-word sets, {minha}, {terra}, and {tem}, respectively, and the fourth entry is the set of all nouns in the Portuguese language (and possibly beyond). Both of the lines, "Minha terra tem saudade" (Lopes, 1963: 8), from Fernando Lopes' 1963 "Minha terra só tem tanques," a foreshadowing of the military coup, and "Minha terra tem cadeiras" (Lélio, 1884: 1), another nineteenth-century political satire, are instances of this simple template where the only variable is the last position, in which a single noun replaces "palmeiras." By design, every instance of this template contains the 3-gram "minha terra tem," which appears in 61% of the 500 texts. This paper's algorithm finds 315 instances of the T_1 template that do not fit the T_0 template spread among 183 texts (37%) of the 500; in total 53% of texts employ some form of the T_1 template.

Contrast the preceding examples with the line "Minha terra tem criança de rua" from "Minha Terra," a poem from 2008 written about Brazil's homelessness, hunger, poverty and other social ills (Bispo, 2008). This line contains six words as opposed to the original's four, and subsequently the template definition must be expanded to remove the limit of four words in its instances. This can be accomplished by allowing any Noun Phrase (rather than a single noun) to occupy the fourth position in the template, creating an ordered list in which the last entry is a set whose elements are sequences of words:

$$T_2 = [\{minha\}, \{terra\}, \{tem\}, Noun\ Phrase]$$

It is easy to see that the template T_2 is a generalization of T_1, as every noun is also a noun phrase. Now, the first three slots contain singleton sets of words while the last slot contains the syntactic category Noun Phrase. Syntactic categories are more general than parts-of-speech: a part-of-speech is a property of a single word, such as noun or verb, whereas a syntactic category can also include sequences of words. In the original poem, the second line, "Onde canta o sabiá," can be interpreted as modifying "palmeiras," creating a single noun phrase, fitting the first two lines of "Canção do Exílio" neatly into the T_2 template.

In "Minha Terra," Cesar Miranda's enumerative montage from 2004, the first two lines read "Minha terra tem cachoeiras / Cascatas, cinismos e falsidades." All four of the nouns "cachoeiras," "cascatas," "cinismos," and "falsidades" are objects of the verb "ter," and it makes little sense to only count the first, "cachoeiras." Rather than make a distinction between conjoined noun phrases and noun phrases headed by a single noun, the T_2 template includes the possibility that the last position contains a collection of noun phrases. In total, the algorithm finds 105 instances of the T_2 template that are not instances of the T_1 template; together there are 420 instances of the T_2 template with a noun or noun phrase in the last slot different from "palmeiras." These 420 instances are found in 225 (45%) of the 500 texts; many texts employ this template more than once.

Considering lines such as "Minha terra não tem recursos" from "Uma Canção" (Camila Vanessa et. al., 2006), the existence of the negation "não" requires yet another shift. The template

$$[\{minha\}, \{terra\}, Adverb, \{tem\}, Noun\ Phrase]$$

is sufficient to generate this line as an instance, as well as the previously mentioned "Minha terra só tem tanques" (Lopes, 1963), however at first glance the template appears to no longer generate the original line "Minha terra tem palmeiras," since no adverb appears before the verb "tem." To overcome this difficulty, the third entry in this template could contain the empty set (a null choice), effectively making this third slot optional, whereas the first, second, fourth and fifth slots do not have this alternative. Conceptually, the template is

$$T_3 = [\{minha\}, \{terra\}, Adverb\ or\ empty, \{tem\}, Noun\ Phrase]$$

With this modification, it is clear that every instance of the template T_2 is also an instance of the template T_3, so the nesting property of the templates is preserved. The algorithm locates 44 instances of this template among 21

texts in which the third slot is not empty. In 36 of those instances (82%), the adverb is "não," indicating this template is predominantly used to expose what "minha terra" lacks, as in Mario Quintana's well-known text from 1962, "Minha terra não tem palmeiras / E em vez de um mero sabiá, / Cantam aves invisíveis / Nas palmeiras que não há" (Quintana, 2006: 443).

Next, consider the line "Minha rua tem crianças" from an untitled poem written in 2013 (Santos et. al.). Here the noun "rua" has substituted "terra," but it is easy to imagine that a noun different from "terra" in an instance of this template would require the entry in the "minha" slot to agree in gender and number with it, as in the line "Nosso país tem coisas feias" from the same poem (Santos et al, 2013). However, if the noun in the second slot (the substitution for "terra") is plural, the verb "tem" would need to change to "têm," and consequently the fourth slot must allow for either the choice of "tem" or "têm." Putting these observations together into one whole, the template becomes

[Possessive Determiner, Noun, Adverb or empty, {tem, têm}, Noun Phrase]

It is noteworthy that instances of this template also occur in the original poem itself, for example, in the line "Nossas várzeas têm mais flores."

Further considering possibilities for the determiner in the first slot, possessive adjectives can shift the ownership of the subject of the instance from the first person (my or our) to the second or third (your, his, her, or their). Yet, the original's first-person perspective is a major feature of the text and the corpus. As Enslen argues, "the autobiographical nature of the poem's responses has been a constant from the nineteenth century to today, as those who write variations continue to work out their place within the nation" (Enslen, 2022: 202). However, at times this shift does occur. An example can be found in the lyrics of Brazil's national anthem (1909), "Do que a terra, mais garrida, / Teus risonhos, lindos campos têm mais flores; / 'Nossos bosques têm mais vida', / 'Nossa vida' no teu seio 'mais amores'," which employs a second-person possessive for "campos," referring back to "terra" and quoting the original. Despite such exceptions, and although a human reader might be able to locate more general instances of this template where the determiner need not be first-person, to avoid this difficulty entirely, the template is restricted in the first slot to only first-person determiners; that is, the set {meu, minha, meus, minhas, nosso, nossa, nossos, nossas}, rendering the template

$$T_4 = [\text{1st Person Determiner, Noun, Adverb or empty,} \{\text{tem, têm}\}, \text{Noun Phrase}]$$

This template finds robust representation in the data set of 500 texts; in total, the algorithm encountered 171 instances of this template that are not instances of the T_3 template, distributed among 89 texts.

In one more twist on the T_4 template, in the original poem, the eighth line, "Nossa vida mais amores" appears without the verb "tem," although "tem" is implied, since the three preceding lines are all instances of the T_4 template. Strictly speaking, "Nossa vida mais amores" is not an instance of the T_4 template; however, it is obvious to the human reader that, in reality, it is related. To handle this case, the template could be modified to add an empty choice to the fourth position:

$$[\text{1st Person Determiner, Noun, Adverb or empty,}$$
$$\{\text{tem, têm, empty}\}, \text{Noun Phrase}]$$

The difficulty with this modification is that it casts a wide net and captures sequences of words that are not genuinely reliant upon the first line of "Canção do Exílio." In the data set of 500 poems, the algorithm counts 280 instances of this template with no verb, but a preliminary reading finds only 32, or 11%, are meaningful recapitulations of "Minha terra tem palmeiras." Consider "meu pulmão as armas," from "Minha terra tem poluição" by Claudio Fernando (2012); although this phrase fits the template, it does not express implied ownership in the same manner as "Nossa vida mais amores."[5] Still, there are texts employing this modified template in an essential way. In "Canção do Exílio" by Luiz Martins (1950), clearly a related text due to the title, the lines "Nossos dias têm mais sol, / nossas noites mais luar, / nossas ruas têm mais gente, / nossa gente mais amores" invoke the template twice.[6] Also, the poem "Canção da Corrupção," by Pamela Berger (2013), contains the lines "Nossas ruas tem [sic] mais buracos, / Nossos bairros só poeira." Despite these interesting adaptations, due to the small percentage of truly relevant uses, this study defines the T_4 template without the empty choice in the fourth slot, creating yet another situation where an algorithmic reading could differ from a close reading.

An equally subtle change appears in the line, "Minha terra tinha palmeiras," from Cleiton Cunha's 2012 nostalgic pastiche, which is identical to the original's first line except in verb tense. To move past the present tense, the template for this line will include as an entry a lexeme, which is a set of words related through inflection (Matthews, 1991). From each lexeme

---

5   When a text does not have a formal title, the first line is used instead, as is the case with this text and others. These make-shift titles are placed in ["Brackets"] in the References section.

6   Capitalization follows the original.

one distinguished element, its lemma, is chosen. For example, the words "tenho", "tinha", "teve" are all elements of the lexeme whose lemma is TER. From the line "Minha terra tinha palmeiras" and the line "Minha terra tinha cachoeiras" from the same poem, the template

$$[\{minha\}, \{terra\}, \text{TER}, Noun]$$

is formed, wherein the third slot the lexeme TER appears. Matching phrases based on shared lexemes has been used previously to establish intertextuality (Forstall et.al, 2015). It is easy to imagine that the flexibility conveyed in the previously discussed templates could also be employed, resulting in the template

$$T_5 = [\text{1st Person Determiner, Noun, Adverb or empty, TER, Noun Phrase}]$$

Interestingly, this template allows for the creation of instances sharing no words, verbatim, in common with the original first line; however, in lemma form, there will be overlap, as both "tem" and any element of the lexeme TER share the same lemma. The algorithm identifies 23 instances of the T_5 template that do not fit the T_4 template; in 19 of the 23 the form of the verb is "tinha," and all but one of those 19 begin with "minha terra tinha" or "meu país tinha." The list of things following these phrases reflects the paradisiacal imagery of the original: arara, um rio, prazeres, palmeiras (6), primores (2), árvores (3), aves, frutíferas árvores, and cachoeiras.

But, is it possible to use a verb other than "ter" in a syntactic template? If so, then even in lemma form, there could potentially be no textual overlap between an instance of such a template with the original first line. The lines "Nossa terra quer sossego" and "Nossa terra não quer freios" (Lopes, 1963) come close to satisfying this condition, where the transitive verb "querer" has replaced "ter," although there is still textual overlap with the original due to the presence of the word "terra." This leads to the possible template

$$[\text{1st Person Determiner, Noun, Adverb or empty,} \\ \text{Transitive Verb, Noun Phrase}]$$

At a minimum, the algorithm could require the verb to be transitive, that is, possessing a direct object, due to the original's reliance on nouns as the "primary descriptive variables" (Enslen, 2022: 16).[7] Thus, a line such as "Minha

---

7   Aurélio Buarque de Holanda, in 1944, first noted the preponderance of nouns and the lack of "qualificativos." He writes, "A falta desse elemento decorativo [qualificativos] valoriza de maneira singular os substantivos do poema, dando-lhes relevo, dilatando-

terra é muito rica" from "Canção" (François Seul, 1881), while expressing some overlap with the original line, "Minha terra tem palmeiras," is not of the same syntactic structure since the phrase "muito rica" is not a noun phrase. Additionally, this template is quite broad and instances of it could readily appear in texts that have no significant relationship to "Canção do Exílio." For example, the phrase "meu coração abriga o infinito," from Adriano Nunes' "Era para Ser uma Canção de Exílio" is also an instance of this template. In many of these instances, the tie is so tenuous that more evidence within the text is required to establish intertextuality. In the Nunes example, the title holds evidence as well as the appearance of Significant Words.[8] The poet writes, "nunca vi sabiás / nem saberia dizer como é um. / palmeiras? só as conheço / de longe. / mas o meu coração abriga o / infinito" (Nunes, 2012). In Seul's nineteenth-century political parody, if the title does not hold sufficient evidence, then the remainder of the poem confirms the relationship through lines such as: "Minha terra tem empregos," "Em cismar que dia e noite," and "E assim permita Deus / Que ainda venha a ser ministro, / Deputado ou senador" (Seul, 1881).

One potential solution to this problem is to make a distinction between an instance (of any type) and an instance that is relevant to the universe of "Canção do Exílio." That is, the template is

$$T_6 = [\text{1st Person Det, Noun, Adverb or empty, Transitive Verb, Noun Phrase}]$$

but to be considered relevant, instances are subject to certain restrictions based on evidence of intertextuality stemming from other categories of analysis, such as Significant Words. For example, in a T_6 instance where the transitive verb is not a form of "ter," if at least one of the words "terra," "palmeiras" or "sabiá" appears, then it could be considered a relevant instance; one example is "minha terra possui aves" from "Belezas de Minha Terra" (Silva, 1860: 3-4). It might be possible to reach the same conclusion if closely related words to "terra," such as "país" occur in the instance; such as the line "meu país canta o melro" from "Uma Resposta," an incisive Portuguese counterargument to G. Dias' Brazil-focused text (R Carlos, 1853). Additionally, as the phrase "nossos céus roubaram estrelas" in "Canção para o Exílio" (Silva, 2010) demonstrates, one of the other Significant Words found in the original, such as "Deus" or "céu," is likely enough to make the required connection.

-lhes a sugestão emocional" (33).

8   The significant words are terra, palmeiras, sabiá, Deus, céu, amores, vida, noite, and flores.

Yet, proof of relevance need not be confined to the words that form the instance. Evidence may instead be located within a text but outside of the words that comprise the instance. The T_6 instance "nossa praça faz calor" in "Gentes dos Cajuais Alerta!!! Canção Popular" (O Muribeca, 1862: 3) is somewhat related to "Minha terra tem palmeiras," but stronger evidence appears in the first stanza, which employs a T_2 template: "Minha terra tem um foro / Onde gira um cajuá / Animal tão bruto assim / Não tem aqui como lá." The algorithm uncovers 47 instances of the T_6 but not the T_5 template, and 27, or 57%, of those instances contain at least one Significant Word. This proportion seems high enough to reasonably include this template in analysis, although close readings may reveal that some are not true reinterpretations of the first line of the original poem.

Conceptually, the category of all possible reworkings of the first line of "Canção do Exílio" are placed in the "T-infinity" category, which includes among many others the consciously omitted phrases akin to "Nossa vida mais amores." Likewise, the line "Minha terra sem palmeiras" (Silva 2010) is a T-infinity instance; this phrase does not fit any of the templates T_0 through T_6, yet only differs from the first line of the original poem by one letter. The templates defined in this study capture many, but not all of the ways "Minha terra tem palmeiras" has been modified in the past, and the ways it will be reimagined in the future.

As this type of template analysis may be useful outside of the context of "Canção do Exílio," in clear and precise terms, a *Syntactic Template* is an ordered list

$$T = (x_1, x_2, \ldots, x_n)$$

whose entries, the x_i's, are sets of sequences of words (including sequences of length 1). Examples of these sets include syntactic categories, lexemes, single words, or other predetermined collections of words relevant to analysis. An *Instance* of a Syntactic Template is a sequence of sequences of words

$$S = (s_1, s_2, \ldots, s_n)$$

where for each i between 1 and n, s_i is an element of x_i. The sequence of sequences of words in an instance are then placed end-to-end to create one contiguous sequence of words that reads naturally as a phrase.

To put the theory into practice, an algorithm was developed to detect instances of the templates described in this paper using a Natural Language Processor (NLP).[9] As remarked earlier, the T_0 template, or any template all

---

9 This study used the spaCy NLP https://spacy.io/

of whose n-many entries are single words, in practical terms is no different than an n-gram. However, the templates T_2 through T_6 are distinct from POS-n-grams and POS-n-skip-grams, since in these templates the number of words in their instances is not fixed in advance while, in contrast, each POS-n-(skip)-gram contains exactly n-many words. Furthermore, in templates T_1 to T_5 there is less choice than there would be in a POS-n-(skip)-gram, as at the very least the verb is required to be a form of "ter." Syntactic templates are at times more restrictive than POS-n-(skip)-grams, and at others less so, and therefore are neither a specific case nor a generalization of that concept, but a separate idea altogether. The algorithm to detect templates relies on the NLP to generate part-of-speech tags, noun phrases, and dependency relationships, and uses this information to search for patterns of words fitting the templates.

The total number of each type of instance found by the algorithm in this study is provided in Table 1. The notation $T_i \backslash T_{i-1}$ refers to instances that are T_i but not T_{i-1} (for i between 1 and 6). For example, $T_3 \backslash T_2$ is used to signify instances that are T_3 but not T_2. Looking at the fourth row of the table (i = 3), there are 21 texts that contain at least one T_3 instance that is not a T_2, 314 texts that contain any T_3 instance (including instances of T_0, T_1, and T_2), 44 instances of T_3 templates that are not instances of T_2 (and so are also not T_0 or T_1), and a total of 618 instances of the T_3 template. This division aids in the classification of instances more similar to the original, as, for example, "minha terra tem primores" is a T_1 instance that is not T_0, but is also a T_2 through T_6 instance; however, it does not employ all of the machinery of the latter templates.

i	Texts with $T_i \backslash T_{i-1}$	Texts with T_i	$T_i \backslash T_{i-1}$ Instances	T_i Instances
0	106	106	154[10]	154
1	183	263	315	469
2	72	302	105	574
3	21	314	44	618
4	89	342	171	789
5	13	347	23	812
6	39	356	47	859

---

10 There are 155 occurrences of the 4-gram "minha terra tem palmeiras," but only 154 instances of the T_0 template due to differences in the algorithms for counting each. The poem "Uma Canção do Exílio" (Barros, 2009) contains the phrase "minha terra tem. Palmeiras," with punctuation. The n-gram calculator ignores punctuation whereas the language processor does not.

An inspection of Table 1 reveals that a total of 356 texts, or 71% of the 500, employ at least one of the templates T_0 to T_6. There are a total of 705 phrases among the 500 texts that are modifications, to some degree, of the original phrase, "Minha terra tem palmeiras." The degree of modification can be estimated by locating the smallest i (between 0 and 6) for which the phrase in question is an instance of T_i; a smaller i indicates a stronger tie.

The rather large numbers of instances in this table and their thematic variability are evidence to support the theory of Syntactic Templates as generators of intertextuality. These templates are indeed being put into practice. Yet, some nuance is lost in the T_0 to T_6 hierarchy, as there are instances more closely related to the original line than their template category might indicate, such as "Minha terra tinha palmeiras," a T_5 instance mentioned earlier. This line differs only by one word, sidestepping most of the complexity built into the template. A different arrangement of the template hierarchy, such as permitting a change in tense in an earlier template, would mark this phrase as more related. What's more, some phrases, such as "Minha terra sem palmeiras" are missed entirely. In this garden of forking paths, the choice of any one in particular will inevitably preclude others.

It should be noted that, while NLPs are continuing to improve, the data they generate are not perfect, especially for an underserved language such as Portuguese. In this study, there were several cases of sentences parsed incorrectly, words tagged with an incorrect part-of-speech, and noun phrases inaccurately extracted. There are additional challenges to parsing language written in verse, since poetic license allows for loose applications of grammatical rules, as illustrated by the original's eighth line "Nossa vida mais amores." In this light, the numbers in Table 1 are estimates of the true prevalence of these templates.

A template is a set of instructions for generating text (its instances). That is, a template is not a collection or sequence of words, and templates cannot be seen in the text of "Canção do Exílio" alone. It is in looking at the various ways in which the first line, "Minha terra tem palmeiras," has been reworked throughout history that the nature of its underlying Syntactic Templates becomes apparent. Syntactic Templates live in the space between the text of the original and the text of a variation. With each reading of "Canção do Exílio," the reader creates her own Syntactic Templates, which may then be used to create a variation and thus participate in the ongoing negotiation of Brazilian identity as reflected in the ever changing Song of Exile corpus.

As there are more lines in "Canção do Exílio" than the first one, and subsequent lines also frequently appear reworked in variations, in particular "Onde canta o sabiá" and "Não permita Deus que eu morra," a similar template analysis of these other lines could yield more insights into the

nature of borrowing in the corpus and the different themes represented in the intertexts. Similarly, these operationalized templates can lead to the discovery of new variations. For example, a search for the phrase "Minha terra tinha" should point to different works than a search for "Minha terra tem".

WORKS CITED

Anonymous (1853). "Minha Terra Natal." *Marmota Fluminense*, 25 Nov 1853. 3. http://memoria.bn.br/DocReader/706906/1720 [11 Apr 2023]

Arruda, Leandro (2012). "Minha Terra Tem Facebook." *Canção do Exílio Revisitada: Processos de Intertextualidade-Intertextos.* Jaboatão dos Guararapes, PE: Organização, 2012. 22. www.scribd.com/document/109651993/Cancao-do-Exilio-Revisitada [10 Apr 2023] Link ativo.

Barros, Robertson Frizero (2009). "Uma Canção do Exílio." *Germina: Revista de Literatura e Arte.* www.germinaliteratura.com.br/sabiaseexilios/asavesqueaquigorjeiam_9.htm [10 Apr 2023]

Battles, Paul (2019). "Using N-gram Analysis to Map Intertextual Networks in Old English Verse." *Digital Philology: A Journal of Medieval Cultures* 8, no. 2 (2019): 155-191. doi:10.1353/dph.2019.0022. [11 Apr 2023] Link ativo.

Berger, Pamela (2013). "Canção da Corrupção." *E.E.B. Willy Hering — Rio do Sul.* 3 May 2013. http://eebwillyhering2.blogspot.com/2013/05/cancao-da-corrupcao.html [11 Apr 2023]

Bispo, Lourdes Bernadete C. (2008). "Minha Terra." *Para Ler e Repensar.* http://www.paralerepensar.com.br/lourdes_minhaterra.htm [no longer active]

Briguet, Paulo (2009). "Canção do Exíguo." *Germina: Revista de Literatura e Arte.* www.germinaliteratura.com.br/sabiaseexilios/asavesqueaquigorjeiam_8.htm [10 Apr 2023].

Broder, A. Z. (1997). "On the Resemblance and Containment of Documents." In *Proceedings. Compression and Complexity of SEQUENCES.* Salerno, Italy, 13 June 1997: 21-29. Institute of Electrical and Electronics Engineers. DOI: 10.1109/SEQUEN.1997.666900I [10 Apr 2023]

Camila et al (2006). "Uma Canção." *Germina: Revista de Literatura e Arte.* https://www.germinaliteratura.com.br/sabiaseexilios/asavesqueaquigorjeiam_3.htm [11 Apr 2023]

Carlos, R (1853). "Uma Resposta." *Periódico dos Pobres*, 27 Oct 1853. Rio de Janeiro. 5. memoria.bn.br/DocReader/709697/2073

Carvalho, Miriam Panighel (2008). "Paráfrase da Canção do Exílio de Gonçalves Dias." *Recanto das Letras.* 27 Mar 2008. https://www.recantodasletras.com.br/poesias-patrioticas/918426 [10 Apr 2023]

Cristina, Samara and Breno Lobato (2013). "[Minha Terra tem o Brega]." *Silene Faro: Língua e Literatura.* infolaboratorio.blogspot.com/2013/03/parodia-da-cancao-do-exilio.html Link ativo.

Cunha, Cleiton (2012). "[Em Minha Terra Tinha Palmeiras]." *Canção do Exílio Revisitada: Processos de Intertextualidade-Intertextos.* Jaboatão dos Guararapes, PE: Organização, 2012.

16. www.scribd.com/document/109651993/Cancao-do-Exilio-Revisitada [10 Apr 2023] Link ativo.

Elagina, Daria (2022). "Encoding of Text Reuse in the Project Beta maṣāḥǝft." *Journal of the Text Encoding Initiative.* Rolling Issue. 1-20. https://doi.org/10.4000/jtei.3763

Enslen, Joshua Alma (2022). *Song of Exile: A Cultural History of Brazil's Most Popular Poem, 1846–2018.* Purdue Studies in Romance Literatures. https://docs.lib.purdue.edu/psrl/1 Link ativo.

Fernando, Claudio (2012). "[Minha Terra Tem Poluição]." *Canção do Exílio Revisitada: Processos de Intertextualidade-Intertextos.* Jaboatão dos Guararapes, PE: Organização, 2012. 19. www.scribd.com/document/109651993/Cancao-do-Exilio-Revisitada [10 Apr 2023] Link ativo.

Ferreira, Luiz Alexandre Cruz (2015). "Ode ao Exílio." *O Alvoradense*, 6 Jan. 2016. www.oalvoradense.com.br [link no longer active] Link ativo.

Forstall, Christopher, et al (2015). "Modeling the Scholars: Detecting Intertextuality through Enhanced Word-level N-gram Matching." *Digital Scholarship in the Humanities* 30.4. 503-515.

Gonçalves Dias, Antônio (1846). "Canção do Exílio." *Primeiros Cantos: Poesias.* Rio de Janeiro: E. and H. Laemmert, 1846. 9-10. books.google.com/ books?id=ma5OAQAAIAAJ

Guthrie, David, et al (2006). "A Closer Look at Skip-gram Modelling." *LREC.* Vol. 6. Link ativo.

Ganascia, Jean-Gabriel, Peirre Glaudes, and Andrea Del Lungo (2014). "Automatic Detection of Reuses and Citations in Literary Texts." *Literary and Linguistic Computing* 29.3 (2014): 412-421.

Kristeva, Julia (1986). "Word, Dialogue and Novel." *The Kristeva Reader.* Ed. Toril Moi. New York: Columbia UP, 34–61.

Lee, John (2007). "A Computational Model of Text Reuse in Ancient Literary Texts." In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, edited by Annie Zaenen and Antal van den Bosch, 472–479. Prague: Association for Computational Linguistics. https://aclanthology.org/P07-1060.pdf. Link ativo.

Lélio (1884). "[Minha terra tem cadeiras]." *Jornal do Recife*, n. 00245, 23 Out 1884. 1. http://memoria.bn.br/DocReader/705110/22012 [11 Apr 2023]

Lioma, C. & van Rijsbergen, C. (2008). "Part of speech n-grams and Information Retrieval." *Revue Française de Linguistique Appliquée*, XIII, 9-22.

Lopes, Fernando (1963). "Minha Terra só Tem Tanques e que Ameaçam Atirar." *Tribuna da Imprensa*, 7 Oct. 1963. 8. memoria.bn.br/DocReader/154083_02/13813 [11 Apr 2023]

Martins, Luiz (1950). "[Minha Terra Tem Palmeiras]." *Estado de São Paulo*, 14 Mar 1950. 6. https://acervo.estadao.com.br/pagina/#!/19500314-22954-nac-0006-999-6-not [11 Apr 2023]

Matthews, Peter H (1991). *Morphology.* Cambridge University Press.

Miranda, Cesar (2004). "Minha Terra." *Germina: Revista de Literatura e Arte.* https://www.germinaliteratura.com.br/sabiaseexilios/asavesqueaquigorjeiam_3.htm [11 Apr 2023] Link ativo.

Muribeca (1862). "Gentes dos Cajuais Alerta!!!" *Pedro II*, 6 Oct. 1862. Fortaleza. 3. memoria.bn.br/DocReader/216828/6335 [10 Apr 2023] Link ativo.

Nunes, Adriano (2012). "Era para Ser uma Canção do Exílio." *Que faço como que não faço.* 12 Mar 2012. https://astripasdoverso.blogspot.com/2012/03/adriano-nunes-era-para-ser-uma-cancao.html. [10 Apr 2023]

Pokou, Yao Jean Marc, Philippe Fournier-Viger, and Chadia Moghrabi (2016). "Authorship Attribution Using Small Sets of Frequent Part-of-Speech Skip-grams." In *Flairs conference*, 86-91. http://www.philippe-fournier-viger.com/FLAIRS2016__AUTHORSHIP_ATTRIBUTION.pdf

Portela, Manuel (2022). "Literary Research as Data Art: An Experiment in Critical Reading." Afterword for *Song of Exile: A Cultural History of Brazil's Most Popular Poem, 1846–2018.* 157-63. Purdue Studies in Romance Literatures. https://docs.lib.purdue.edu/psrl/1 Link ativo.

Quintana, Mário (2006). "Uma canção." *Poesia Completa.* Org. Tânia Franco Carvalhal. Rio de Janeiro: Editora Nova Aguilar, 443.

Santos, Maria Riciene N. and Janeth Lucia (2013). "[Minha Rua Tem Crianças]." *Silene Faro: Língua e Literatura.* 18 Mar 2013. http://infolaboratorio.blogspot.com/2013/03/parodia-da-cancao-do-exilio.html [14 Apr 2023]

Seul, François (1881). "Canção." *O Aspirante.* 17 Dez 1881. 4. http://memoria.bn.br/DocReader/232017/44 [11 Apr 2023]

Silva, João Nepomuceno da (1860). "Belezas de minha terra." *Gazeta Oficial*, vol. 3, no. 90., 21 Apr 1860. Belém do Pará. 3-4. memoria.bn.br/DocReader/720836/1507 [10Apr 2023] Link ativo.

Silva, Sérgio (2010). "Canção para o Exílio." *Penetra Surdamente no Reino das Palavras.* 31 Jan 2010. http://consideracaodopoema.blogspot.pt/2010/01/cancao-para-o-exilio.html. [10 Apr 2023].

Soares, Jô (1992). "Canção do Exílio às Avessas." *Veja*, 16 Sep 1992. 15.