



Universidade Federal do Rio de Janeiro

Cibele Ribeiro da Cunha Oliveira

**UM EXPERIMENTO PARA A AUTOMAÇÃO
DO PROCESSO DE AVALIAÇÃO TEXTUAL
ATRAVÉS DA PROSPECÇÃO DE PADRÕES
GRAMATICAIIS EM NARRATIVAS**

DISSERTAÇÃO DE MESTRADO



Instituto de Matemática



Instituto Tércio Pacitti de Aplicações
e Pesquisas Computacionais

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
INSTITUTO TÉRCIO PACITTI DE APLICAÇÕES E PESQUISAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Cibele Ribeiro da Cunha Oliveira

UM EXPERIMENTO PARA A AUTOMAÇÃO DO
PROCESSO DE AVALIAÇÃO TEXTUAL ATRAVÉS DA
PROSPECÇÃO DE PADRÕES GRAMATICAIS EM
NARRATIVAS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Informática.

Orientador: Prof.^a Claudia Lage Rebello da Motta, D.Sc., UFRJ

Rio de Janeiro

2019

Oliveira, Cibele Ribeiro da Cunha

Um Experimento para a Automação do Processo de Avaliação Textual Através da Prospecção de Padrões Gramaticais em Narrativas / Cibele Ribeiro da Cunha Oliveira. – 2019.

Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, Programa de Pós-Graduação em Informática, Rio de Janeiro, 2019.

Orientadora: Claudia Lage Rebello da Motta

1. Machine Learning (Teses). – 2. Processamento de Linguagem Natural (Teses). – 3. Avaliação Textual (Teses). – 4. Símbolos Gramaticais (Teses). – I. Claudia Lage Rebello da Motta (Orient.). II. Universidade Federal do Rio de Janeiro, Instituto de Matemática Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, III. Título.

CIBELE RIBEIRO DA CUNHA OLIVEIRA

**UM EXPERIMENTO PARA A AUTOMAÇÃO DO
PROCESSO DE AVALIAÇÃO TEXTUAL ATRAVÉS
DA PROSPECÇÃO DE PADRÕES GRAMATICAIS EM
NARRATIVAS**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Informática.

Aprovada em 20 de novembro de 2019.

Prof.^a Claudia Lage Rebello da Motta, D.Sc., UFRJ

Prof. Claudio Miceli de Farias, D.Sc., UFRJ

Prof. Sean Wolfgang Matsui Siqueira, D.Sc., PUC

Prof. ^a Simone Ballmann de Campos, D.Sc., UFSC

*Dedico esse trabalho, ao meu filho,
Miguel Roman de Oliveira Marque,
Não importa o quão difícil
pareça para o mundo,
sonhe alto e sonhe muito.
Persista, resista e não desista.*

Agradecimentos

Obrigada à Professora Claudia Motta que acreditou no meu trabalho desde o princípio e me apoiou em cada etapa sempre me incentivando a desenvolver minha pesquisa da forma máxima e objetiva possível, confiando em sua viabilidade e colaboração com a sociedade.

Ao professor Carlo Emmanoel que sempre acreditou no meu trabalho e me deu toudo o apoio necessário a programação para que o Spekuloom fosse concebido ,

À professora Carla Veronica a quem devo grande parte e meu conhecimento sobre a como a educação deve ser feita.

A Bruna Fiuza que foi minha companheira de mestrado desde o inicio dessa empreitada, compartilhando as alegrias e tristezas ao longo de todo o processo.

À Ana Paula Cavadas que me apoiou na escolha de fazer o mestrado e me mostrou o que e batalhar para realizar um sonho.

À Isabel Hortência que sempre topou todas os desafios da academia, estando presente e colaborando com todos os seus processos.

Ao Daniel Chaves que foi um exemplo de como trazer a abstração de conceitos complicados em formato de um discurso passível de entendimento.

À Emanuelle e Dayara que sem a colaboração de vocês nos processos de digitalização e debate sobre o tema nada seria possível.

À Leniah que sempre esteve ao meu lado nos momentos de maior necessidade acadêmica, seja para a entrega de artigos ou para realização de oficinas.

À minha mãe Josane, pelo apoio incondicional e motivação que me deram base para querer construir um futuro melhor.

Ao meu pai e mestre Carlo Emmanoel, por acreditar em minhas capacidades e me estimular a ousar acima de tudo.

Às minhas irmãs Ana Luiza e Beatriz, por serem o meu porto seguro emocional.

Ao, pai do meu filho, Bruno, que compreendeu a importância de meus estudos e se disponibilizou para estar ao lado de Miguel em qualquer momento que eu precisasse me dedicar ao aprendizado.

Ao meu filho Miguel, que a cada dia, sem sombra de dúvida, me faz querer ser uma pessoa melhor.

Provavelmente não sou um romancista; provavelmente eu sou um ensaísta que precisa de escrever romances porque não sabe escrever ensaios.

José Saramago

Resumo

Ribeiro da Cunha Oliveira, Cibele, 2019. Dissertação de Mestrado em Informática – Instituto de Matemática, Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.

O analfabetismo funcional é o principal vetor do fracasso escolar, visto que a dificuldade na interpretação ocasiona quebra de comunicação, rompendo com o aprendizado. Recuperar a interlocução significa voltar-se para o letramento, e a escrita, por sua vez, é parte integrante deste processo na escolaridade. Para observar a produção escrita do aluno, os professores propõem a elaboração de textos adequada para os diversos níveis de conhecimento. A escrita apresenta estruturas de coesão e coerência textual que sinalizam em que etapa da aquisição da escrita do ponto de vista do letramento a criança está. A redação, por sua vez, é corrigida em grandes quantidades pelos professores que, devido à demanda, podem deixar de avaliar questões importantes pela falta de tempo ou atenção. O Spekuloom é um modelo computacional que captura dados para a investigação de estruturas textuais e das construções gramaticais próprias da narrativa, tem como um de seus objetivos apoiar a avaliação feita pelo professor sinalizando quais estruturas foram utilizadas pelos alunos e qual o nível de desenvolvimento textual eles se encontram. O estudo tem a metodologia de Maria Montessori como precursora da análise gramatical feita através da marcação simbólica. Ao tratar da linguagem, faz a análise morfológica através dos símbolos gramaticais e suas famílias, catalisando a compreensão e criação do texto por meio do domínio das regras de construção linguística. A computação identifica, mas ainda não analisa as estruturas recursivas gramaticais, fazendo-se necessário o seu estudo para melhor entender as competências linguísticas textuais. Esta é a proposta do Spekuloom: classificar textos em diferentes gradações com base na recursividade de padrões gramaticais ao submetê-los à inteligência artificial.

Palavras-chave: avaliação textual, símbolos gramaticais, *machine learning*, educação, processamento de linguagem natural.

Abstract

Ribeiro da Cunha Oliveira, Cibele. **Um Experimento para a Automação do Processo de Avaliação Textual Através da Prospecção de Padrões Gramaticais em Narrativas**. 2019. Master Thesis in Informatics – Instituto de Matemática, Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.

Functional illiteracy is the main vector of school failure, since difficulty in interpretation causes communication breakdown, disrupting learning. Recovering the conversation means turning to literacy, and writing, in turn, is an integral part of this schooling process. Teachers propose the elaboration of texts suitable for the different levels of knowledge to observe the student's written production. Writing presents structures of cohesion and textual coherence that indicate at what stage of the writing acquisition from the point of view of literacy the child is. The essay, in turn, is corrected in large quantities by teachers who, due to demand, may fail to evaluate important issues for lack of time or attention. Spekuloom is a computational model that captures data for the investigation of the own narrative textual structures and grammatical constructions. One of its objectives is to support the assessment made by the teacher by signalling which structures were used by the students and what level of textual development they have achieved. The study has the Maria Montessori Methodology as a precursor of the grammatical analysis made through symbolic marking. When dealing with language, it makes the morphological analysis through the grammatical symbols and their families, catalyzing the comprehension and creation of the text through the mastery of the linguistic construction rules. Computation identifies, but does not yet analyze grammatical recursive structures, its study is necessary to better understand textual language skills. The goal of Spekuloom proposal is to classify texts into different gradations based on the recursion of grammatical patterns by subjecting them to artificial intelligence.

Keywords: text evaluation, grammar symbols, machine learning, education, NLTK.

Lista de Figuras

Figura 1. Períodos Sensíveis Montessori.....	23
Figura 2. Encaixa de Ferro	28
Figura 3. Símbolos Gramaticais Montessori	30
Figura 4 . Exemplo de Arquitetura Usada em NLTK para Diálogos Falados.....	33
Figura 5. Exemplo de Análise Sintagmática Feita pelo NLTK.....	33
Figura 6 . Tópicos para a Análise Textual de um Parágrafo	37
Figura 7. Software Montessori que Transforma Material Concreto em Digital.....	40
Figura 8. Processo de Tokenização do NLTK.....	41
Figura 9. Análise Gramatical Simbólica Feita pela Máquina.....	42
Figura 10. Exemplos de Padrões e sua Recursividade	42
Figura 11. Textos da Categoria Básica.....	43
Figura 12. Laboratório Orange	43
Figura 13. Padrões e sua Recursividade Comparando Níveis Básico e Intermediário.....	44
Figura 14. Comparação entre Avaliações Textuais após a Análise Bayesiana	44
Figura 15. Configuração do Aprendizado de Máquina do Laboratório Orange.....	48
Figura 16. Matriz de Confusão para a Classificação Naive Bayes.....	48
Figura 17. Família do Nome.....	51
Figura 18. Família do Verbo.....	51
Figura 19. Análise de Padrões Gramaticais de Dez Textos Submetidos.....	54
Figura 20. Análise de Padrões Gramaticais de Trinta e Quatro Textos Submetidos.....	55
Figura 21. Avaliação Textual Subdividido em Básico, Intermediário e Transitório.....	56
Figura 22. Padrões Recursivos em Grupos de Quatro Palavras	60
Figura 23. <i>Scatter Plot</i> de Padrões em Diferentes Categorias.....	60
Figura 24. <i>Scatter Plot</i> da Avaliação Textual dos Alunos separados por Gênero: Feminino e Masculino	61
Figura 25. <i>Scatter Plot</i> da Quantidade de Padrões de Diferentes Níveis Presente no Texto de Cada Aluno	61

Lista de Tabelas

Tabela 1 . PISA 2015 – Parâmetro para Avaliar Proficiência em Leitura.....	70
Tabela 2. Enem Competência 1.....	71
Tabela 3. Enem Competência 2.....	72
Tabela 4. Enem Competência 3.....	73
Tabela 5. Enem Competência 4.....	74
Tabela 6. Enem Competência 5.....	75
Tabela 7. Primeira Tabela da Documentação de Avaliação Textual de Padrões	98
Tabela 8. Segunda Tabela da Documentação de Avaliação Textual de Padrões	100
Tabela 9. Terceira Tabela da Documentação de Avaliação Textual de Padrões.....	102

Sumário

INTRODUÇÃO	14
1.1 PREÂMBULO	15
1.2 CARACTERIZAÇÃO DO PROBLEMA	18
1.3 HIPÓTESE	19
1.4 METODOLOGIA DE PESQUISA	20
1.5 PERGUNTA DA PESQUISA	20
1.6 OBJETIVO	20
1.6.1 Objetivo Geral	21
1.6.2 Objetivo Específico	21
1.7 RELEVÂNCIA DO ESTUDO	21
1.8 ORGANIZAÇÃO DO ESTUDO	21
2 REFERENCIAL TEÓRICO.....	22
2.1 DESDE A AQUISIÇÃO DA FALA AO LETRAMENTO EM CONTEXTO ESCOLAR	23
2.2 AVALIAÇÃO DA ESCRITA NO CENÁRIO EDUCACIONAL	24
2.3 EDUCAÇÃO MONTESSORI	26
2.3.1 Os Símbolos Gramaticais Montessori	29
2.4 TECNOLOGIA NA EDUCAÇÃO	31
2.5 O PAPEL DA TECNOLOGIA NA AVALIAÇÃO.....	32
2.6 PROCESSAMENTO DE LINGUAGEM NATURAL.....	32
2.7 TRABALHOS RELACIONADOS.....	34
2.7.1 Símbolos Gramaticais na Tecnologia em prol do Aprendizado da Língua Inglesa como uma Segunda Língua.	34
2.7.2 Avaliação da Riqueza Textual.....	34
2.7.3 Letramento e Tecnologia.....	35
3 PROPOSTA	36
3.1 O SURGIMENTO DO MODELO VOLTADO PARA O AUXÍLIO À AVALIAÇÃO TEXTUAL.....	38
3.2 CRIAÇÃO DO SOFTWARE SPEKULOOM	39
4 METODOLOGIA	46
4.1 TIPO DE PESQUISA.....	47
4.2 SELEÇÃO	47
4.2.1 Seleção Textual	47

4.2.2 Seleção de Indivíduos.....	47
4.3 DESCRIÇÃO DO INSTRUMENTO UTILIZADO	47
4.4 ANÁLISE NO LABORATÓRIO ORANGE.....	48
5 REALIZAÇÃO DO EXPERIMENTO Implementação do Spekuloom.....	49
5.1 COLETA DE INFORMAÇÕES	50
5.2 DEFINIÇÃO DE OBJETIVOS.....	50
5.3 PROJEÇÃO DO EXPERIMENTO.....	52
5.4 OPERAÇÃO DO EXPERIMENTO	53
5.5 ANÁLISE DO EXPERIMENTO.....	54
5.6 INTERPRETAÇÃO DOS RESULTADOS	54
5.7 ADEQUAÇÃO DO EXPERIMENTO.....	55
5.8 OPERAÇÃO DE CONFIRMAÇÃO DO EXPERIMENTO	55
5.9 APLICAÇÃO DOS RESULTADOS	56
6 ANÁLISE DE RESULTADOS	58
6.1 AVALIAÇÃO TEXTUAL E RESULTADOS	59
7 CONCLUSÃO.....	63
Referências	67
Anexos 69	
anexo a – Descrição e percentual de estudantes nos sete níveis de proficiência em leitura – PISA 2015.....	70
Anexo b – Avaliação da redação do Enem – competência 1.....	71
Anexo c – Avaliação da redação do Enem – competência 2.....	72
anexo d – Avaliação da redação do Enem – competência 3.....	73
anexo e – Avaliação da redação do Enem – competência 4.....	74
anexo f – Avaliação da redação do Enem – competência 5	75
anexo g	75
Apêndices 76	
Apêndice a – Documentação primeira parte.....	76
Apêndice B – documentação segunda parte	99

INTRODUÇÃO

“Quando a criança começa a pensar e a usar a linguagem escrita para expressar seus pensamentos básicos ela está pronta para o trabalho elementar, esta aptidão é uma questão não da idade ou outra circunstância, mas de maturidade mental”.

Maria Montessori

O presente capítulo trata de uma breve introdução do assunto de pesquisa abordado nesta dissertação. Consta entre seus tópicos a caracterização do problema, a hipótese, a metodologia de pesquisa, a pergunta da pesquisa, os objetivos gerais e específicos, a relevância do estudo e sua organização.

1.1 PREÂMBULO

O analfabetismo funcional, ou seja, os analfabetos e os alfabetizados em nível rudimentar; é tido como um importante fator no fracasso escolar. Em uma escala de abaixo de 1b, sendo este o menor nível, até 6, sendo este o maior; mais do que a metade da população brasileira chega apenas até o nível 1a de proficiência em leitura de acordo com Fernandes(2016) e com base no PISA 2015. O que significa que mais de 50% da população apenas localiza fragmentos, reconhece assuntos principais e faz correlações simples. A dificuldade na interpretação de uma questão leva a uma quebra de comunicação que culmina no rompimento do processo de aprendizado.

A língua é um dos núcleos principais educativos, pois é através dela que ocorre a interlocução. A quebra de comunicação interrompe a mensagem que o emissor leva ao receptor, comprometendo o processo de aquisição do conhecimento. A língua como instituição social Saussure (1907), existe apenas em seres concretos na sua coletividade. É por meio dela que é possível o ensinar e aprender. Mas se, ao longo do processo, algo ocorre e corta o envio/recepção da mensagem – algum ruído que pode estar no ambiente físico ou no campo da língua – o aprendizado se rompe e a informação é perdida. Para recuperar essa informação, é importante que, de algum modo, essa mensagem seja enviada.

Observar a língua de modo geral é entender que as palavras possuem classes gramaticais nas quais dentro de cada grupo podem substituir umas as outras sem ferir a coesão da frase, salvo exceções. Saber que cada classe de palavra tem combinações possíveis finitas em uma frase – vertente da sintaxe – e o conhecimento sobre esse afinilamento das possibilidades de agrupamento de classes gramaticais permite um maior domínio do letramento da leitura e da escrita, permitindo o acesso à mensagem recebida ou enviada.

Enquanto o domínio do código da leitura e escrita é nomeado alfabetização, de outro modo, a capacidade de um aluno utilizar a linguagem como prática social em diversos contextos, dominando a leitura e a escrita é chamada de letramento. É através do letramento que o aluno consegue interagir em inúmeras circunstâncias sem que haja ruptura na comunicação. A alfabetização é o ato de decifrar o código da linguagem escrita através do conhecimento das letras, dos fonemas a ela relacionados e das convenções de pontuação. Por outro lado, o letramento é o entendimento e utilização correta da língua nos diversos

ambientes sociais. A alfabetização e o letramento deveriam caminhar juntos na evolução do aprendizado da língua pelo aluno, pois um é complemento do outro nas interações escrita/leitura e fala/compreensão oral.

Decifrar o código da escrita é um processo mais rápido do que o processo do letramento. O último requer que o aluno esteja em contato com as diversas formas de comunicação escrita e oral para que saiba utilizá-las nos diferentes tipos de contexto. O processo de desenvolvimento do letramento carece de um apoio que vai além da leitura e escrita no contexto social propriamente dito, falta àquele alguma forma de materialização de sua representação para que haja um modelo a ser seguido pela criança, servindo-lhe de base metacognitiva para seu aprendizado. É no letramento, tanto na prática de leitura quanto de escrita, que se observa uma estrutura linguística que sustenta a língua por trás da semântica das palavras, ou seja, há uma trama que segue uma ordem para que no texto haja coesão e coerência.

Se por um lado a alfabetização e o letramento são processos de aprendizagem que ocorrem na escola desde a primeira idade, por outro o aprendizado da sintaxe de uma língua requer uma longa dedicação dos estudos desta e, nem sempre, o educando é capaz de transferir o que aprendeu pelo canal de entrada, leitura, para o canal de saída, sua produção escrita. A semântica e a sintaxe dentro dos critérios de coesão e coerência são a essência da interpretação textual. Esta, por ser um tópico significativo para a compreensão e realização de tarefas tanto de estudo quanto de trabalho, é parte integrante de concursos públicos, sendo critério de avaliação presente em todas as suas provas.

Então, o que pode ser feito para diminuir a lacuna entre a dificuldade de aprimoramento do letramento do educando sendo este critério avaliativo para o ingresso em uma universidade ou carreira?

O papel do educador como mediador deste processo é fundamental, pois este como especialista de sua área deve ser capaz de avaliar em que etapa seu aluno está e apresentar estratégias que possibilitem a evolução deste para os próximos níveis de proficiência. Na prática, o que acontece nas salas de aula é a produção escrita de redações que o professor corrige uma a uma, muitas vezes deixando de sinalizar pontos importantes devido ao tempo disponível ou desatenção pela quantidade. Para reverter este quadro, é necessário um mecanismo que apoie o professor nesta avaliação, otimizando sua correção e contribuindo, de modo indireto, com a evolução do nível de proficiência do educando.

Sugerir uma abordagem metalinguística como possível mecanismo, no qual se lança mão de uma verificação com estrutura matemática para que seja uma validação da análise do professor sobre nível do texto produzido pelo aluno é abrir caminho para uma mediação textual com métrica, tornando parte do letramento algo perceptível estruturalmente tanto para o professor quanto para o aluno. Este conceito segue a pedra angular do Método Montessori – concretizar conceitos abstratos – e trilha uma solução para a evolução do indivíduo observável pela escrita, respeitando os estágios do desenvolvimento da criança Piaget (1975). Tornar parte do conceito de letramento em algo visualmente tangível abre espaço para uma avaliação textual por parte do professor comprometida com essa métrica e para um modelo evolutivo padrão no qual o aluno projeta sua escrita de um modo cada vez mais adequado ao propósito social em que está inserido.

Criar uma ferramenta computacional que possibilita a formação desse retrato do letramento da escrita, captura momentânea do nível desta, corrobora de forma indireta para que não somente diminua o Analfabetismo Funcional, como também expanda a interlocução entre o aluno e as diversas áreas do conhecimento através de uma maior compreensão da língua. Através da programação dos algoritmos apropriados é possível desenvolver um engenho que busque por padrões e avalie o texto com base nestes, tornando-se um apoio à correção textual do professor. Este, por sua vez, observa os padrões e consegue visualizar de forma concreta como auxiliar o seu aluno a desenvolver melhor seu texto. É possível, também, que o aluno, de modo autônomo, possa ver seus resultados e com base nestes desenvolver uma estratégia de aprimoramento de sua escrita.

A escrita representa a tradução das ideias em palavras, quanto maior a coesão e coerência do texto maior o entendimento de quem o lê. A leitura e a escrita são vinculadas de modo que ao aprimorar uma a outra se desenvolve em paralelo. Logo, uma ferramenta que facilita o avanço da escrita, conseqüentemente é responsável por aperfeiçoar a leitura e vice e versa. Preparar um ambiente com condições propícias para que se desenvolvam a leitura e a escrita em diferentes circunstâncias, significa lançar mão de uma ferramenta que auxilia no processo de letramento, desimpedindo a comunicabilidade.

Não há como falar da leitura e da escrita sem atrelá-los ao ambiente virtual. Com os dispositivos móveis presentes no cotidiano, grande parte da linguagem está diretamente acessível através da tecnologia. Fazer uso desta para alcançar o processo de aprendizagem e pinçar a escolha de padrões gramaticais que estão sendo utilizados faz com que um novo formato avaliativo textual surja. Uma ferramenta computacional que avalie a etapa de

aquisição da escrita do ponto de vista do letramento a partir de um texto tem potencial para auxiliar no desbloqueio da comunicabilidade.

1.2 CARACTERIZAÇÃO DO PROBLEMA

Analisar e Avaliar uma produção textual não é tarefa fácil. Dentro das quatro habilidades inerentes a aprendizagem de uma língua – compreensão oral, fala, leitura e escrita– a escrita é a de nível de dificuldade mais elevada, no entanto a leitura tem papel fundamental para que se forme um escritor competente, porém é necessário ir além e lançar mão de estratégias diretamente relacionadas ao texto produzido para que a proficiência de seu autor se desenvolva.

O aprendizado de uma língua envolve diferentes etapas. Entre elas, encontram-se a aquisição do vocabulário, a estruturação sintática de frases, a conjugação verbal, entre outras. Essas etapas estão vinculadas à vivência da criança na sua primeira infância em sociedade. A linguagem então surge aproximadamente aos dois anos de idade juntamente com a estruturação simbólica iniciando o processo de elaboração das significações Piaget (1975). Além do aprendizado ligado à oralidade, há a representação simbólica que é relativa à alfabetização, representação dos fonemas da língua através das letras do alfabeto, esse aprendizado ocorre aproximadamente entre as idades de quatro e oito anos, no contexto brasileiro. Porém, em paralelo ao aprendizado da língua, mencionado, ocorre o letramento que nada mais é do que dominar a língua de modo que o indivíduo saiba utilizá-la nos diferentes contextos. Este processo normalmente se dá através da exposição da criança à leitura, mas esta, por si só, não basta para que na hora de realizar uma produção textual o escritor tenha propriedade sobre a língua.

A questão principal tratada neste trabalho, a avaliação de uma produção textual, está diretamente ligada a relevância de se apropriar da escrita questão ressaltada Soares. Contudo, apesar da leitura propriamente dita ser de grande valia, não garante que o processo de letramento se desenvolva na criança a ponto desta ser capaz de construir um texto tão elaborado quanto o que é capaz de ler. O que se observa, com frequência, é que apesar da formação do leitor ter um papel ativo no despertar da leitura, nem sempre caracteriza o processo de formação dos mesmos em escritores competentes.

A questão é construir uma educação que, de maneira eficaz, leve à criação de indivíduos efetivamente letrados, ou seja, que dominem não só as nuances gramaticais e sintáticas da linguagem, mas consigam se expressar dentro de todo o espaço de significância da língua.

Aprender a língua como código de comunicação em qualquer de suas habilidades é uma tarefa que se inicia com a compreensão oral e culmina na escrita. Nos primeiros meses de vida, o bebê logo consegue reconhecer algumas palavras da fala e até interagir com elas. A compreensão oral da língua é um processo cognitivo passivo, ou seja, é um canal de entrada desta. Porém, é a partir dela que após alguns anos a criança começa a experimentar e a produzir sons. O mesmo ocorre com o processo de leitura e escrita. Ler é um canal de entrada da língua enquanto que a escrita é um processo de produção da mesma. Sendo a leitura primordial para que as estruturas de linguagem escrita, diferentes da oral, comecem a pertencer ao imaginário esquemático do aprendiz.

Por ser o topo da pirâmide, no que diz respeito à apropriação da língua, escrever na configuração formal dentro dos diferentes contextos sociais requer uma dedicação intensa e enfrentamento dos inúmeros erros e acertos até que um texto esteja propriamente escrito dentro das normas de coesão e coerência, utilizando-se adequadamente os esquemas sintáticos. Por isso, é de um elevado nível de dificuldade para a o aprendiz.

O problema levantado por este estudo é a dificuldade do professor em auxiliar o aluno, de modo concreto, a realizar a apropriação sociocognitiva da língua escrita no que diz respeito à configuração formal e ao seu uso social. O que ocorre, em sua grande maioria, é a proposta de um tema de redação pelo professor para que o aluno infira, com sua leitura e conhecimentos prévios, qual contextualização e formato apropriados para tal tarefa. Após o feito, entrega para o professor para que esse corrija os erros e sugira caminhos mais adequados a serem tomados. A correção destes textos requer muita atenção para que não se deixe passar questões importantes, porém o tempo e a quantidade de textos a serem corrigidos são os maiores obstáculos desses profissionais. Corrigir uma redação é uma atribuição feita ainda de modo bastante subjetivo e passível de falhas¹ o que é reafirmado tanto pelos concursos públicos que montam uma equipe de corretores de redação quanto pela avaliação de trabalhos acadêmicos que passam pelas mãos de pelo menos dois avaliadores.

1.3 HIPÓTESE

Para que o professor seja capaz de auxiliar de modo mais preciso no processo de letramento do aluno a ponto deste se apropriar da escrita assim como se apropriar da leitura, é necessário lançar mão da vertente matemática da língua. A hipótese deste estudo é levantar a

¹ Notícia G1 sobre correção de redações por profissionais contratados para a equipe do Enem. Disponível em: <<https://g1.globo.com/educacao/enem/2016/noticia/corretores-de-redacao-do-enem-avaliam-em-media-74-redacoes-por-dia.ghtml>>

viabilidade do uso das classes gramaticais e de parte da sintaxe como uma forma mensurável de classificação do texto em um nível de metalinguagem, ou seja, uma linguagem utilizada para falar sobre outra linguagem. Essa abordagem permite que se analise um texto, não pelo seu significado semântico, mas pela classificação gramatical das palavras utilizadas neste, o que inaugura a hipótese de que é possível analisar e por consequência avaliar um texto através da recursividade de sequências de classes gramaticais presentes neste. A avaliação, então mencionada, seria uma forma de medir o nível de letramento da escrita presente em um texto.

1.4 METODOLOGIA DE PESQUISA

A metodologia adotada para a pesquisa é quantitativa, pois por meio de geração de dados estes podem ser transformados em estatísticas na avaliação de cada texto submetido à ferramenta computacional Spekuloom.

1.5 PERGUNTA DA PESQUISA

Partindo da hipótese de que é possível avaliar o nível de letramento da escrita de um texto através de sequências de classes gramaticais que se repetem ao longo deste, as principais perguntas a serem tratadas nesta pesquisa são: Há padrões gramaticais em sequência que se repetem em diferentes tipos de texto? Caso haja, é possível utilizar esses padrões para avaliar se um texto é de nível básico, intermediário ou transitório? É possível criar uma ferramenta computacional, que baseada na estruturação do textual a partir das sequências de classes gramaticais, avalie o texto a ela submetido, apoiando o professor ou aluno autônomo de modo a descobrir em que nível de letramento da escrita aquele se encontra, possibilitando um novo parâmetro para projetar estratégias inovadoras de melhoria da escrita?

1.6 OBJETIVO

Os objetivos deste estudo estão correlacionados de modo indireto com a necessidade de dar ao letramento a mesma ênfase que é dada à alfabetização na língua. Para que isso seja possível, a ponto de ser aplicado como uma prática, é imprescindível o uso da computação como meio pelo qual se torna viável a classificação textual em massa através das classes gramaticais em um curto espaço de tempo. A computação pode tornar o que seria um trabalho enfadonho do ponto de vista humano, numa simples etapa de todo o processo de letramento.

1.6.1 Objetivo Geral

Este trabalho tem como objetivo apresentar uma ferramenta de avaliação textual com foco na coesão gramatical, ou seja, todos os processos de sequencialização que asseguram (ou tornam recuperável) uma ligação linguística significativa entre os elementos que ocorrem na superfície textual de acordo com Mateus (1983). Esta ferramenta tem o intuito de matematizar parte da avaliação textual a tornando mais objetiva.

1.6.2 Objetivo Específico

Quanto aos objetivos específicos, pode se dizer que são os seguintes:

- Criar uma ferramenta computacional que seja capaz classificar textos a partir de suas classes gramaticais.
- Ensinar para esta ferramenta os diferentes níveis de competência textual, ou seja, o quanto o texto analisado possui de apropriação da língua. Níveis estes que são baseados na recursividade de sequências de classes gramaticais.
- Fazer com que a partir deste ensinamento, a ferramenta seja capaz de analisar novos textos e categorizá-los no nível adequado.

1.7 RELEVÂNCIA DO ESTUDO

Engendrar uma ferramenta computacional educacional que não seja apenas um substituto do material didático direcionado ao letramento, como por exemplo, livros online, já marca um diferencial nesta pesquisa. Porém, usar a mesma tecnologia como meio de aperfeiçoar o processo avaliativo tanto em relação à otimização do tempo utilizado, quanto em relação à acurácia da correção dos textos, em si, torna clara a relevância deste estudo.

1.8 ORGANIZAÇÃO DO ESTUDO

O presente estudo apresenta o processo de criação de um Framework Linguístico chamado Spekuloom como ferramenta computacional previamente mencionada e que será desenvolvido em quatro partes: na primeira os preâmbulos, na segunda o referencial teórico e trabalhos relacionados, na terceira, apontamos à proposta da pesquisa realizada, na quarta, a metodologia utilizada, na quinta a realização do experimento, na sexta os resultados da análise e na sétima as conclusões apresentadas na pesquisa.

2 REFERENCIAL TEÓRICO

Ter-se apropriado da escrita é diferente de ter aprendido a ler e a escrever: aprender a ler e escrever significa adquirir uma tecnologia, a de codificar em língua escrita e de decodificar a língua escrita; apropriar-se da escrita é tornar a escrita “própria”, ou seja, assumi-la como sua “propriedade”.

Magda Soares

Neste capítulo, serão apresentadas a conceituação de letramento e alfabetização no contexto escolar. Em seguida, serão tratadas diferentes formas de avaliação do letramento em leitura e letramento em escrita. Logo depois, haverá um breve relato sobre a educação Montessori e como esta trata o assunto. Haverá, também, uma seção sobre o uso da tecnologia na educação. Finalmente, abordará o estado da arte em tecnologia avaliativa, uma introdução sobre NLTK e fechará com alguns trabalhos relacionados.

2.1 DESDE A AQUISIÇÃO DA FALA AO LETRAMENTO EM CONTEXTO ESCOLAR

O aprendizado de uma língua é caracterizado por quatro habilidades as quais começamos a desenvolver desde o nascimento. Começa quando o bebê através da audição processa os sons e passa a processar sua função simbólica por volta dos dois anos de idade Piaget (1964) e, por conseguinte, entendem que a junção destes criam palavras que possuem significado. Após a iniciação da compreensão oral, o bebê passa a testar os sons até que eles façam sentido para seus ouvintes, surgindo, assim, a fala. Em seguida, entra a escola com a função de tornar esse indivíduo alfabético e letrado.

A alfabetização é distinta do letramento, porém ambas são indissociáveis. A alfabetização é um processo antigo de aquisição do sistema convencional da escrita, que possui diferentes vertentes e métodos, no entanto, o seu cerne é alfabético e ortográfico. O termo letramento, por sua vez, surgiu há aproximadamente três décadas atrás, em função da urgência tanto na vida social quanto em atividades profissionais de situações centradas e dependentes da língua escrita. O termo foi criado para suprir a necessidade de denominar o desenvolvimento de comportamentos e habilidades do uso competente da leitura e escrita em práticas sociais.

Segundo Montessori (1987), a criança de zero a seis anos possui o que ela denomina de *Mente Absorvente*. É uma etapa de intenso desenvolvimento cerebral que Montessori dividiu em *Períodos Sensíveis* representados na Figura 1.

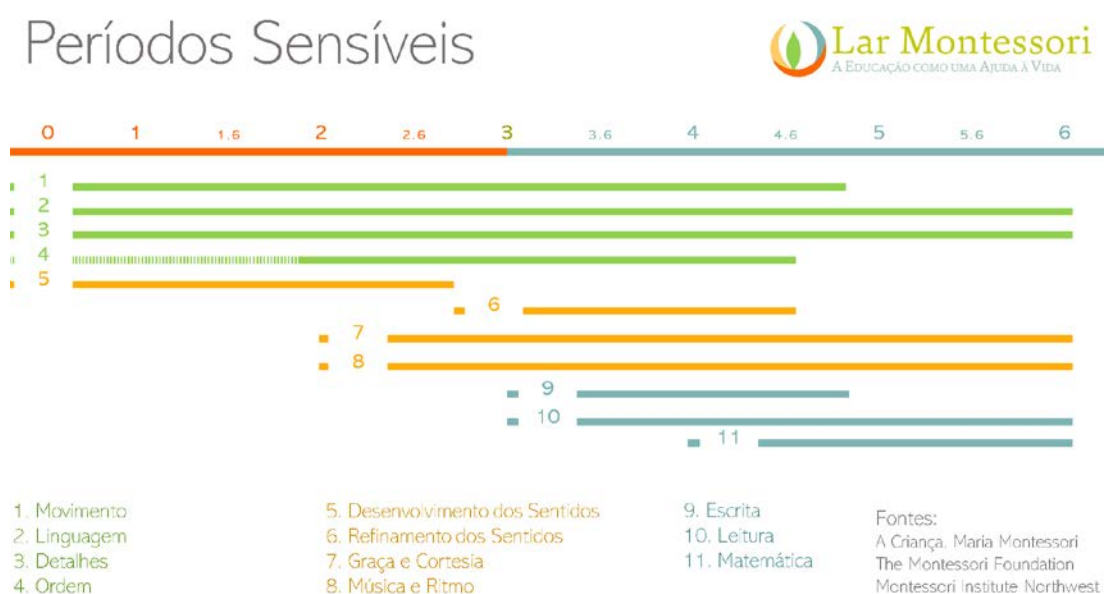


Figura 1. Períodos Sensíveis Montessori

Nestes períodos, observa-se que o processo de escrita e leitura se inicia por volta dos três anos, isto significa que a atenção da criança passa para estes dois processos de modo a estar pronta para adquirir e desenvolvê-los. É interessante notar que o período sensível da escrita se inicia aos três anos e se finaliza por volta de cinco, entre os quais há o fenômeno da explosão da escrita que de acordo com sua metodologia Montessori (1949) ocorre por volta dos quatro anos se a criança passar pelos estímulos apropriados. O que significa que a alfabetização é um meio de adquirir o sistema-código da língua, deste modo, é adquirido em um tempo relativamente curto de aprendizagem. É importante frisar que a escrita nesta imagem para Montessori é caracterizada pela alfabetização. Por outro lado, a leitura é um processo contínuo que se inicia aos três anos de idade e continua por tempo indeterminado. Assim, o letramento não é mais considerado um conjunto de habilidades adquiridas apenas na infância, mas um conjunto crescente de conhecimentos, habilidades e estratégias construídos ao longo da vida em ambiente social mais ampla do que a alfabetização e que a engloba de acordo com Tfouni (2002):

“[...] a necessidade de se começar a falar em letramento surgiu, creio eu, da tomada de consciência que se deu, principalmente entre os linguistas, de que havia alguma coisa além da alfabetização, que era mais ampla e determinante desta”.

Soares (2005) reforça esse conceito, sinalizando que o aprendizado da alfabetização é algo que vai desde a parte motora como a forma correta de segurar o lápis até a parte cognitiva como, por exemplo, a sequenciação textual. Por outro lado, o letramento é a ampliação progressiva do termo alfabetização para o conhecimento da leitura e da escrita para exercer uma prática social.

2.2 AVALIAÇÃO DA ESCRITA NO CENÁRIO EDUCACIONAL

Há extensa literatura sobre o letramento associado à alfabetização na primeira infância, todavia o presente trabalho não se detém apenas a essa faixa etária, mas permeia toda a educação básica podendo incluir até trabalhos acadêmicos. Porém, como é feita, hoje, a avaliação do nível de letramento na educação básica como um todo? Pode-se apontar para o Programa Internacional de Avaliação Estudantil (PISA)² como parâmetro de avaliação de letramento em leitura que é dividido em três características: situação, texto e aspecto. Tais aspectos serão analisados de modo breve com o intuito de exemplificar o nível de subjetividade empregado nesta com base em Fernandes (2016).

² Foi tomado como base para esta pesquisa o PISA realizado em 2015.

A primeira característica é a situação que se subdivide em: pessoal (satisfaz os interesses do indivíduo), pública (textos relacionados a questões da sociedade em geral), educacional (este se destina à instrução) e ocupacional (leitura que envolve a realização de alguma tarefa direta). A segunda é o texto dividido em duas classificações: formato - contínuo (organizados em parágrafos), não contínuo (compostos por uma série de dados matriz), combinados (mistura dos dois) ou múltiplos (independência dos dois tipos, mas que se complementam em um significado) - e tipo - descrição (refere-se a propriedades de objetos no espaço), narração (refere-se a propriedade do objeto no tempo), exposição (informações como conceitos compostos nos quais podem ser analisados), argumentação (relaciona conceitos e proposições), instrução (orientação sobre o que fazer) ou interação (procura alcançar finalidade específica) . Finalmente, há o aspecto que na prova se divide em três: localizar e recuperar (transita entre descrever como chegar a informação necessária e selecionar a informação necessária), integrar e interpretar (transita entre compreender a coerência do texto e conectar fragmentos para que tenha um significado) e refletir e analisar (transita entre comparar, contrastar e levantar hipóteses com seu próprio conhecimento e relacionar informações em um texto com conhecimento de fontes externas).

Os resultados dessa análise são reportados em uma escala de proficiência disposta em anexo. Observe o nível de proficiência dos estudantes brasileiros em 2015, os números confirmam o baixo índice de letramento.

Quanto ao letramento em escrita podemos observar outro exemplo que mais uma vez o nível de subjetividade empregado nesta, a tabela de competências abordadas na correção de redações do Exame Nacional do Ensino Médio (Enem) ³. Somam um total de cinco competências: Competência 1 – demonstrar domínio da modalidade escrita formal da língua portuguesa (convenções da escrita, convenções gramaticais, escolha de registro e escolha vocabular), Competência 2 – compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa (competência que avalia as habilidades integradas de leitura e de escrita), Competência 3 – selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista (a elaboração de um texto que apresente, claramente, uma ideia a ser defendida e os argumentos que justifiquem a posição assumida), Competência 4 – demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação (dizem respeito à estruturação

³ Usou-se para este estudo a cartilha do participante do Enem do ano de 2019

lógica e formal entre as partes da redação) e Competência 5 – elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos (sugerir uma iniciativa que busque, mesmo que minimamente, enfrentá-lo).

A avaliação leva em conta os quesitos apresentados e a pontuação é dada de acordo com as tabelas apresentadas em anexo. Ambos os exemplos exprimem o nível elevado de subjetividade para a realização da análise textual. Sem contar com o alto nível de trabalho maçante dos avaliadores que corrigem manualmente questões dissertativas e redações.

2.3 EDUCAÇÃO MONTESSORI

A pesquisa surgiu a partir do estudo original de Dra. Maria Montessori sobre a análise morfológica da língua. Ela elaborou símbolos gramaticais - formas geométricas coloridas que representam as classes gramaticais - e suas famílias - padrões recursivos desses símbolos em textos - para criar um modelo de apoio concreto ao aprendizado de classes gramaticais e sintaxe.

Dra. Maria Montessori elaborou uma metodologia na qual a criança aprende e desenvolve a sua cognição através do uso de materiais concretos os conceitos que são passados de forma abstrata na educação formal básica. A concretização do material tem como fundamento a realização do aprendizado acontecer através das mãos, pois é através do trabalho com estas que a criança chega à concentração espontânea.

A educação é vista de forma diferente em uma escola Montessori. O seu principal pilar é a formação do homem como cidadão do mundo, logo, desde sua primeira infância, as crianças são estimuladas a terem autonomia perante seu aprendizado. Para que essa autonomia seja alcançada, é necessário que os três principais pilares desta metodologia estejam em pleno funcionamento, são eles: o adulto preparado, o ambiente preparado e a criança equilibrada.

A formação de profissionais aptos para a educação Montessori é um processo de longa duração e envolve persistência e resiliência. Os cursos⁴ duram aproximadamente três anos, para que um professor tenha completado o ciclo de estudos básicos referente aos alunos de 3-6 anos de idade por exemplo. Além do curso em si, esses profissionais precisam realizar um estágio de pelo menos cem horas, ler e fichar os livros Montessori e escrever o álbum (escrita sobre a apresentação de materiais Montessori) referente a cada módulo que se dividem em : Conhecimento de Mundo, Linguagem, Matemática, Vida Prática e Sensorial. Ser um

⁴ Exemplo baseado na formação oferecida pelo Curso de Formação de 3-6 na Aldeia Montessori. Disponível em : < <https://www.montessori.com.br/capacitcao-para-professores-montess>>

professor observador em uma sala Montessori requer o domínio da apresentação dos materiais específicos de cada ambiente de sala de aula, além do conhecimento psicopedagógico de cada período sensível da criança a fim de que o professor seja capaz de auxiliar efetivamente cada etapa de desenvolvimento do aluno.

O professor observador é aquele que olha para a sala de aula e percebe qual é a etapa que seu aluno está e quais materiais são apropriados para estimular este aluno a seguir para o próximo nível de desenvolvimento. Para que essa postura do professor seja possível, é necessário que o ambiente esteja preparado para receber este aluno. Um ambiente preparado significa: uma sala com a mobília adequada para a movimentação de criança, estantes com materiais Montessori e folhetos completos dispostos na ordem de dificuldade e professores qualificados contribuir com o processo de aprendizagem da criança.

É natural que a criança, ao se deparar com um ambiente preparado, tenha curiosidade de querer conhecer cada material, logo, ao começar o ano letivo, há um período de adaptação com o ambiente que dura aproximadamente quatro semanas. Nesse período de tempo a criança, através da livre escolha, elege um material para utilizar. Por ela ser dona do seu conhecimento, os materiais eleitos são de seu interesse particular. A estima pelo material faz com que a sua atenção seja cada vez maior a levando a autodisciplina. Nessa metodologia, não é um indivíduo externo que impõe a disciplina, mas a vontade interna de conhecer aquele assunto que faz o aluno fixar o seu foco no material. A autodisciplina é o principal indicativo que a criança está em equilíbrio.

A educação na Metodologia Montessori tem base científica. Quando Dra. Maria Montessori estabelece o papel do professor como um observador, ela o compara a um cientista que observa o seu objeto de estudo Montessori (2014). Neste caso, o objeto de estudo do professor é a forma pela qual uma criança aprende, portanto, o papel deste é seguir a criança e perceber suas etapas de desenvolvimento. A partir da observação, o professor é capaz de sugerir materiais que vão de encontro ao interesse da criança, contribuindo, assim, como sua aquisição do conhecimento. Foi com base nas suas observações sobre a criança que Montessori desenvolveu seus materiais e não o oposto. Logo, ela utilizou de todos os meios e instrumentos disponíveis na época para elaborar técnicas e métodos facilitadores do aprendizado, ou seja, ela utilizou a tecnologia que estava acessível em sua época para criar instrumentos que oportunizam o aprendizado, mostrando que Dra. Maria Montessori usava a tecnologia disponível na época para criar seus próprios materiais, em seu trajeto como pedagoga se esta colaborasse com o avanço cognitivo da criança. Como exemplo pode-se

citar a elaboração do material de Encaixes de Ferro (conjunto de placas de ferro, cada uma contendo uma forma geométrica móvel) representados na figura 2, Montessori (2011) usa a tecnologia, ferramentas e máquinas que transformam matérias primas para se tornarem materiais a partir da aplicação de seu conhecimento técnico e científico sobre este, para elaborar a série de materiais que concretizem as abstrações. Logo, é importante salientar que se a metodologia estivesse sendo elaborada no momento atual, o mesmo estaria acontecendo através da tecnologia vigente.

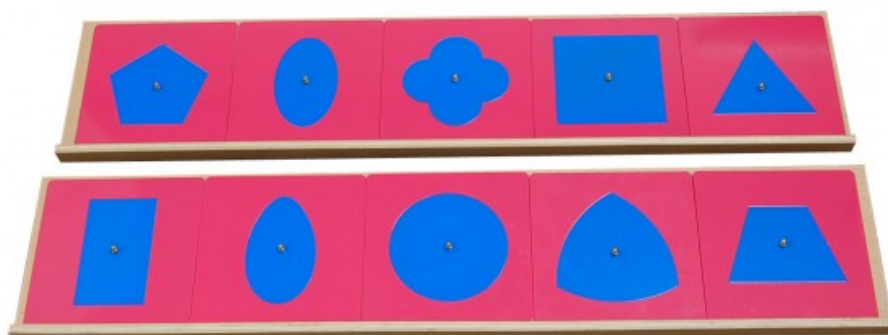


Figura 2. Encaixa de Ferro

A metodologia Montessori trata a educação como um experimento científico, o objetivo é a buscar como ocorre a aprendizagem na criança, retirando os obstáculos de seu caminho. Na pesquisa de Montessori, foram encontrados alguns direcionamentos em relação ao seu objetivo que serão relatados a seguir. Para que o aprendizado aconteça em sala de aula, por exemplo, é fundamental que o professor prepare previamente o ambiente com materiais de desenvolvimento (elaborados por Montessori) para a criança e observe sua movimentação por entre estes. Após a preparação do ambiente, o professor deve ter um olhar apurado para as escolhas de materiais feitas pela criança, pois é ele o responsável por oferecer materiais que desafiem o desenvolvimento desta quando ainda não está pronta para realizar essas escolhas de modo independente. O professor, também, está presente para ajudar em alguma questão quando a criança o solicitar, do contrário, ele observa os erros e acertos desta para que ele possa sugerir outros materiais que reforcem um conteúdo ainda não muito bem assimilado. O educador, deste modo, é responsável por zelar o trabalho - termo utilizado por Montessori para denominar o uso do material em sala de aula - da criança para que esta mantenha a sua concentração.

Ao tratar a linguagem, Montessori quis aplicar a mesma abordagem do uso de materiais concretos no ensino da gramática. A língua é uma simbologia de códigos abstratos e trazê-la para o concreto não é trivial, logo, Montessori criou os símbolos gramaticais em paralelo para que suas formas e cores facilitassem o processo de absorção deste conteúdo.

2.3.1 Os Símbolos Gramaticais Montessori

Na *Psicogrammatica*, Montessori (2017) elabora um estudo de toda a gramática da língua italiana de modo que a criança de seis a doze anos seja capaz, de modo concreto, de realizar a análise morfológica (estudo da estrutura e classificação das palavras em função do uso) do texto, na qual ela simboliza cada classe com uma forma geométrica e uma cor específica. Para cada um das famílias, do nome e do verbo, se traz uma relação de afetividade a denominação (família) para facilitar a sua memorização ao realizar uma atividade relacionada.

Como exemplo de uma apresentação de classe gramatical para as crianças pode-se citar o “Jogo Intuitivo do Verbo” - escrita de bilhetes que são entregues às crianças para que elas realizem o que está escrito neles sem dizer o que é. A criança que viu a mímica tem que adivinhar a palavra, em seguida, ao término da pilha de bilhetes, o professor lê todos em voz alta e diz que são ações e que estas serão representadas por círculos grandes e vermelhos.

A simbolização faz com que o aluno tenha maior facilidade de visualizar as classes gramaticais, tornando-se trivial na metodologia, pois concretiza uma abstração através da abordagem viso motora e áudio fonética para que através de diferentes sentidos seja possível a classificação de modo mais rápido e eficaz. Seguem os símbolos na figura abaixo.

	substantivo
	artigo
	adjetivo
	conjunção
	preposição
	verbo
	adverbio
	pronome
	interjeição

Figura 3. Símbolos Gramaticais Montessori

Mais uma vez a concretização de uma abstração através dos diferentes sentidos, pedra angular da metodologia Montessori, se mostra presente, inclusive, em uma área puramente abstrata que é a língua.

Assim como os símbolos gramaticais em separado, Montessori apresenta, igualmente, agrupamentos de palavras representados pelas duas famílias: a família do nome e a família do verbo. São padrões simples, de três classes gramaticais, os quais a sequência gramatical tem importância. A ideia principal é a referência ao termo sintagma encontrado nos estudos de Saussure (1907), mais especificamente, na divisão entre sintagma nominal e sintagma verbal de um modo apropriado para a idade. A partir da visão Montessori do conteúdo de análise morfológica, introdução da classificação sintagmática e sintática, surgiu a ideia de pesquisar quais eram as possibilidades de padrões em narrativas que fossem recursivas e representassem cada nível de letramento da escrita.

A sintaxe nada mais é do que parte da gramática que estuda a disposição das palavras em uma oração e a relação das próprias orações entre si. Nela, classificam-se os vocábulos em: em sujeito, predicado, objeto (direto e indireto), complemento nominal, aposto, vocativo, adjunto adnominal, adverbial, entre outros elementos. É uma etapa que ocorre posteriormente a apresentação das classes gramaticais. Em paralelo ao estudo da gramática tradicional, há a gramática sintagmática, que sob o ângulo da linguística caracteriza e define de modo mais simplificado do que a primeira.

O presente trabalho propõe encontrar os padrões sequenciais de classes gramaticais canônicos como os da família do nome e do verbo de Montessori e utilizá-los como parâmetro

para avaliação de riqueza textual, ou seja quanto mais rico o texto maior a competência relacionada ao letramento da escrita de seu autor.

Tal tarefa - encontrar padrões canônicos em diferentes narrativas - seria fatigante se feito a olho nu por um professor, de outro modo, empregar a programação como ferramenta de auxílio para, através do conhecimento técnico e científico, constatar esses padrões e realizar essa tarefa em um tempo e esforço menor do que a mão humana é ir de acordo com os princípios científicos de Dra. Maria Montessori e utilizar as tecnologias para desenvolver materiais que concretizam abstrações, ou seja, tornar visível e lógico a metalinguagem por traz de um texto.

2.4 TECNOLOGIA NA EDUCAÇÃO

Acompanhar a evolução tecnológica deveria fazer parte do processo natural da educação. O que se observa é que a tecnologia tem avançado de modo cada vez mais rápido, alvejando um futuro próximo no qual os homens e as máquinas atuarão lado a lado com o intuito de otimizar o tempo gasto em tarefas corriqueiras. A título de exemplo, pode-se citar o desenvolvimento de robôs capazes de realizar, em poucos instantes, tarefas que seriam longas e árduas para qualquer ser humano, através da inteligência artificial. Por outro lado, a área da educação propriamente dita não tem sofrido muitas mudanças se levado em consideração o ambiente (carteiras enfileiradas uma atrás das outra) e o modo de operação das aulas (em sua grande maioria oferecidas de modo expositivo), principalmente quando se refere à educação tradicional. Quando se fala dos recursos de tecnologia disponíveis nas escolas, não se trata apenas de questão ligada à viabilidade financeira, mas da mudança de paradigma que ocorreu na tecnologia e que não foi acompanhada por muitos professores, pedagogos e diretores de estabelecimentos de ensino. O uso do aparato tecnológico em sala de aula traz, a alguns, insegurança devido a diferentes motivos como: pouco domínio do professor e dificuldade de gerência de turma. Ao se tratar de escolas não tradicionais, nas quais há um cuidado maior em relação à filosofia e aos diferentes métodos de aprendizagem, a rejeição se torna ainda maior por muitos acreditarem que a tecnologia traz um embotamento cognitivo devido a diferentes razões⁵. De certo, a tecnologia pode trazer malefícios, mas quando o seu uso é feito de modo inapropriado.

⁵ Exemplo de texto que aborda o tema preocupação com o uso da tecnologia na sala de aula. Disponível em : < <https://larmontessori.com/2013/04/14/e-montessori-a-crianca-e-a-tecnologia/>>

Joly (2002) argumenta que a tecnologia é um recurso estratégico voltado para a ação pedagógica. Valente (2008) coloca que a internet é um bom exemplo da aprendizagem em espiral (ligada ao aspecto social e emocional). Silveira (2001) e Joly discutem a influência da globalização no quadro da educação e aponta os processos de pesquisa e comunicação como diretrizes em oposição a conteúdos fixos, mas apesar da internet ter sido responsável pela abertura desses processos, pode ser subutilizada se certas precauções não forem tomadas. Joly propõe o desenvolvimento de pesquisas sobre leitura e escrita através do computador, criando novas relações entre aluno e professor resultantes da diferente forma de ensinar e aprender.

A relação educação e tecnologia não é um estudo recente, desde os anos oitenta Goodwin et al.(1986) pesquisa sobre os efeitos da utilização de microcomputadores em pré-escolares, existindo uma grande gama de opiniões a favor e especulações contra.

Ramos (2007) demarca o papel da tecnologia em sala de aula como mediadora do processo de comunicação, pois oferece vantagens para a organização, compartilhamento e registro de informações, além da oferta de ferramentas para a comunicação e construção de textos coletivos.

Tratar a tecnologia como ferramenta da sala de aula não é apenas um processo que parte da educação. Uma das iniciativas corporativas é a *Apple Classrooms of Tomorrow* na qual Bertoncello (2010) afirma que sua primeira edição ocorreu entre os anos de 1985 e 1995 e sua segunda edição em 2008. Nesta última, tratou principalmente dos novos discursos e formas de comunicação através das plataformas de aprendizagem virtual que a partir da colaboração e interdisciplinaridade aponta a organização do conhecimento como caminho para a resolução de problemas reais.

Sousa et al. (2016) alega a importância da educação a distância apenas possível através da comunicação existente através da relação aluno-computador o que confirma a colocação de Ramos(2007).

2.5 O PAPEL DA TECNOLOGIA NA AVALIAÇÃO

GRAMARLY

2.6 PROCESSAMENTO DE LINGUAGEM NATURAL

O *Natural Language Toolkit*, mais conhecido como *NLTK*, é uma plataforma com um conjunto de bibliotecas e programas utilizados para o processamento de linguagem natural através de símbolos e estatística Bird (2009). Por ser capaz de classificar as palavras tanto do inglês quanto do português em suas classes gramaticais. O seu uso se torna providencial para

que o processo de espelhamento metalinguístico da *psicogrammatica* Montessori seja realizado, ou seja, a observação não do ponto de vista semântico, mas gramatical das palavras e a transformação dessas classes nos símbolos para que haja melhor manejo quanto à análise morfológica da classificação textual do ponto de vista do usuário.

A seguir a figura 4 exprime colunas com diferentes ramos da linguagem passíveis de serem utilizados no processamento de linguagem em NLTK, porém, para fins da presente dissertação, ir-se-á ater à coluna morfológica.

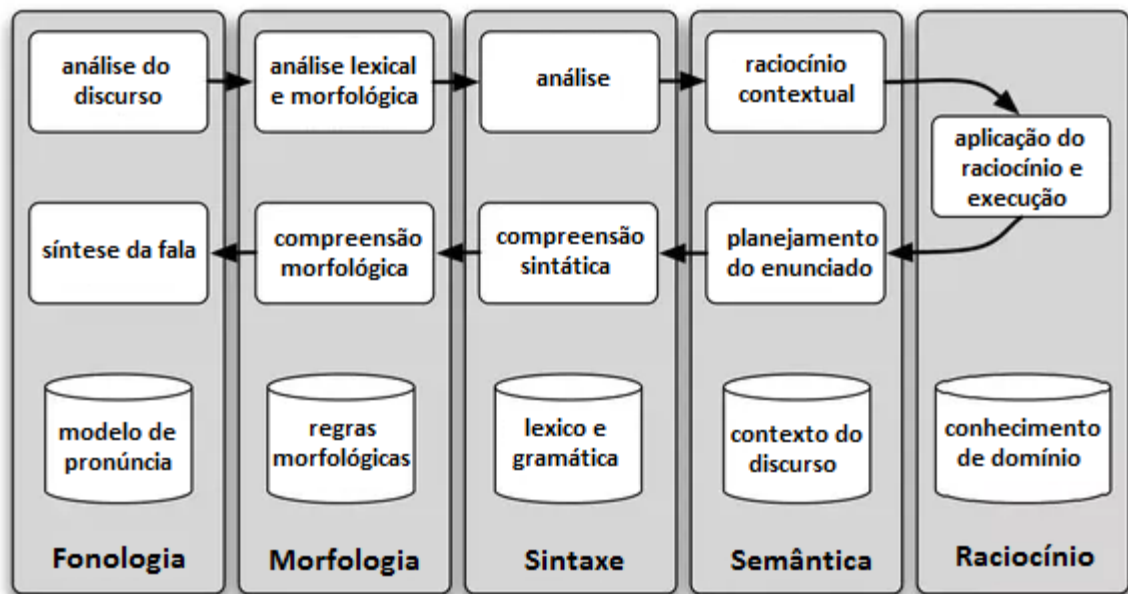


Figura 4 . Exemplo de Arquitetura Usada em NLTK para Diálogos Falados

É interessante, também, colocar que uma das análises possíveis feitas pelo NLTK se aproxima a da gramática linguística separando sintagma nominal de sintagma verbal e classificando seus vocábulos, o que colaborou de modo impar para que a atual pesquisa evoluísse para seu objetivo de separação de padrões.

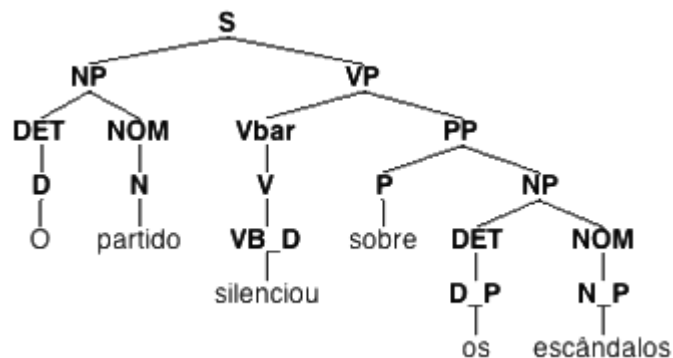


Figura 5. Exemplo de Análise Sintagmática Feita pelo NLTK

2.7 TRABALHOS RELACIONADOS

2.7.1 Símbolos Gramaticais na Tecnologia em prol do Aprendizado da Língua Inglesa como uma Segunda Língua.

Estudos sobre a aquisição da linguagem e letramento são presentes de modo rotineiro na área da educação. Um artigo recente sobre o uso de algoritmos dos Símbolos Gramaticais Montessori Sasi (2018) comprova que o espelhamento simbólico é uma prática relevante quando se trata de simplificar uma abstração para que haja uma concretização desta, promovendo um melhor entendimento da mesma.

O artigo apresenta símbolos gramaticais em inglês desenvolvidos pelo autor para melhorar a compreensão dos alunos e, conseqüentemente, das regras gramaticais relacionadas aos verbos em inglês.

Um experimento foi feito e com duração de trinta horas e este abrangeu dois grupos de alunos um que recebeu aulas de gramática com símbolos gramaticais e outro que recebeu aulas de gramática sem os símbolos.

O teste estatístico revelou maiores pontuações para o grupo que aprendeu gramática com os símbolos. Assim, o autor fortemente recomenda usar estes símbolos, pois a utilização desses permite que os professores e os alunos tenham uma ideia geral sobre o que se esperar em seguida ao os observar.

2.7.2 Avaliação da Riqueza Textual

Quanto à avaliação do texto em si, encontramos um estudo que aborda as diferentes formas de mensurar a riqueza lexical de um texto Casañas (2017), demonstrando que a avaliação textual é um assunto presente na pesquisa científica acadêmica.

Faz uma extensa revisão dos diferentes métodos existente para medir a riqueza lexical de textos. Há uma visão dos principais índices existentes para quantificar a riqueza lexical, explicando como eles são definidos e avaliando seus pontos fortes e fracos.

Propõe uma metodologia de mensuração da riqueza lexical para poder ser utilizada em todo o corpus textual, viabilizando a comparação entre textos e constituindo, assim, uma classificação.

2.7.3 Letramento e Tecnologia

Além da educação, há também estudos na área de informática sobre como melhorar o letramento dos alunos através da utilização de tecnologias como ferramenta otimizadora de atividades que seriam enfadonhas para dar-se apenas pela mão humana Barros (2009).

O artigo evoca a relevância dos gêneros textuais receberem grande atenção nas escolas brasileiras, devido a sua importância no aprendizado da linguagem e, por consequência disso, sendo uma das possíveis maneiras de resolver a questão das altas taxas de analfabetismo funcional.

O trabalho discute formas de implementação de um sistema que analisa características de oito gêneros textuais encontrados em jornais.

O sistema é um jogo e, durante este, o jogador elabora um jornal através da escrita de diferentes gêneros textuais. Esses textos são analisados por um agente Bayesiano que os reconhece.

3 PROPOSTA

La mente umana è come uno specchio dove tutto se riflette e dal quale partono le idee sotto forma di parole.

Maria Montessori

O presente capítulo apresenta a proposta e aborda como o modelo de avaliação textual surgiu e como o software Spekuloom foi elaborado passo a passo para que fosse possível a implementação do modelo.

A ideia dessa pesquisa é engendrar um modelo automático que ao receber os dados da produção textual seja capaz de avaliar em que nível o texto se encontra através da contagem de padrões recursivos gramaticais.

A leitura é fundamental para povoar o imaginário do escritor, no entanto, há poucas ferramentas de apoio para que o aluno perceba qual é o seu nível da escrita e parta deste princípio para desenvolver estratégias de aprimoramento de seus textos.

O caminho mais comum a se percorrer para chegar ao objetivo do letramento com foco na escrita na educação é através de redações escolares, claramente exemplificadas por Kleiman (2005), propostas pelos professores e que, em sua correção, o educador assinala os erros apresentados para que o aluno a refaça do modo adequado ou, ao escrever o próximo texto, leve em consideração o que foi pontuado na escrita anterior. A figura 6 demonstra a subjetividade presente na análise de um parágrafo para caracterizá-lo dentro de um dessas subdivisões, a mesma subjetividade está presente na avaliação de uma redação escolar.

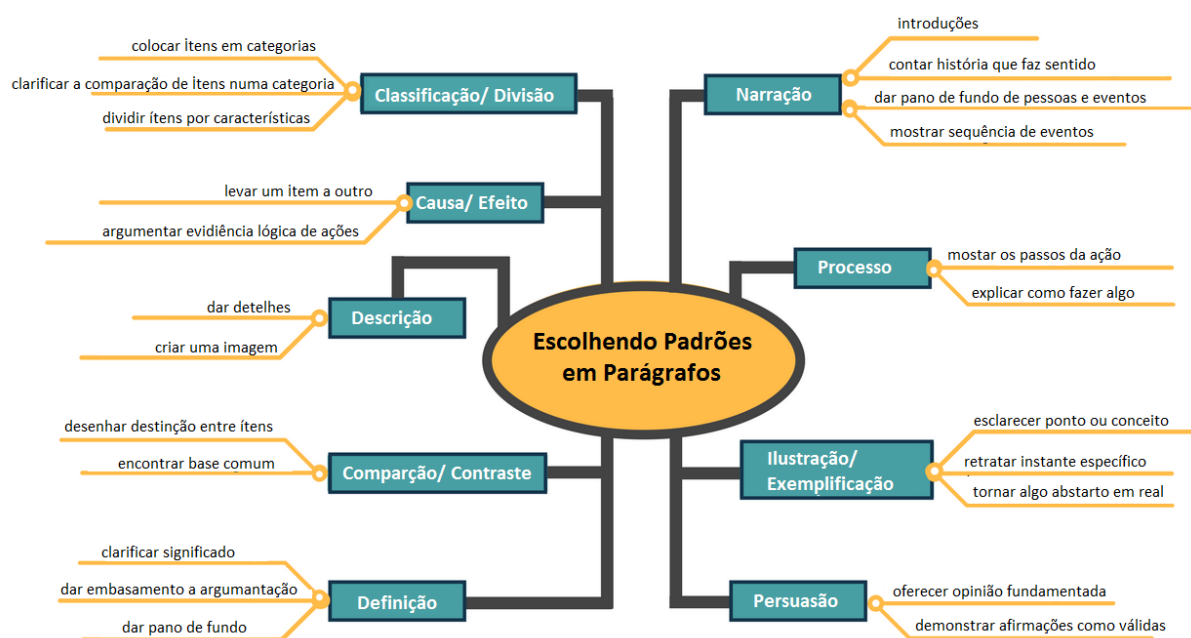


Figura 6 . Tópicos para a Análise Textual de um Parágrafo

A proposta nesta pesquisa é elaborar uma ferramenta capaz de avaliar qual o nível de letramento da escrita de textos narrativos submetidos - o quanto o texto está desenvolvido de acordo com a demanda social por ele apresentado - a partir dos padrões gramaticais sequenciais presentes no próprio, no intuito de localizar em que nível está o autor e iniciar o seu desenvolvimento a partir deste ponto, caracterizando, assim, um olhar individual a cada submissão.

O surgimento da ideia de um modelo que colaborasse com esta parcela do letramento está diretamente ligado ao intuito de unir os três pilares a seguir: Educação, Informática e Sociedade. A busca da convergência das três áreas viabilizou a realização do presente de trabalho.

3.1 O SURGIMENTO DO MODELO VOLTADO PARA O AUXÍLIO À AVALIAÇÃO TEXTUAL

A elaboração do modelo aqui presente mesclou três áreas de conhecimento. Desta união surgiu uma proposta de solução educacional voltada para o âmbito do letramento da escrita narrativa. A sociedade é a primeira área envolvida, com foco na linguagem, no que diz respeito ao letramento que se encaixa na questão social da educação, pois envolve o domínio de uma língua em seus diferentes ambientes sociais. Em seguida, a educação, considerando os estudos da Dra. Maria Montessori, aproveitou-se, no domínio da linguagem, a instrução sobre as classes gramáticas representadas por seus símbolos. Finalmente, a informática participa como instrumento de implementação do modelo com o uso do software Spekuloom na avaliação de textos produzidos por alunos.

Hoje em dia, é dado grande ênfase ao letramento, quando se trata das abordagens de aquisição da linguagem em geral Piaget(1964). Enquanto a alfabetização trata da aquisição progressiva dos símbolos que representam os fonemas do idioma examinado, o letramento envolve, entre outros, a textualidade, ou seja, a coesão e coerência textual Val (1994) entre outros. Por conseguinte, ao apoiar o desenvolvimento do letramento, o presente estudo colabora para que haja uma ferramenta capaz de corroborar com a diminuição gradativa do analfabetismo funcional.

Dra. Maria Montessori, por sua vez, ao longo de seus estudos elaborou uma estratégia ligada à classificação gramatical das palavras Montessori (2016). Ela criou um grupo de símbolos, formas geométricas, e os relacionou a cores específicas, logo o aluno ao analisar um texto de forma gramatical, ao invés de usar palavras como: substantivo e verbo; lançava mão destes símbolos. Esta estratégia otimizava a análise da mesma forma que facilitava a sua apreensão. Como parte desta estratégia, Dra. Maria Montessori também ensinava a família do nome e a família do verbo que nada mais eram do que sequências de classes gramaticais que se repetem em diferentes estruturas textuais. A partir destas sequências que, neste estudo, questionou-se a possível existência de outras sucessões gramaticais.

Seguindo um dos preceitos da Dra. Maria Montessori, no que diz respeito à educação: siga a criança; o presente trabalho se inicia pelo texto produzido por um aluno como ponto de partida para que este avance em sua proficiência, partindo do nível em que se encontra. Quando se analisa os índices de analfabetismo funcional observa-se que a educação carece de abordagens como esta, de acordo com o Relatório do 2º ciclo de monitoramento das metas do Plano Nacional de Educação – 2018, elaborado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), dos adolescentes de 15 a 17:

Com base nas definições adotadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), considera-se alfabetizada a pessoa que declara saber ler e escrever e analfabeta funcional a pessoa com 15 anos ou mais de idade que possui menos de quatro anos de escolaridade ou que declara não saber ler e escrever. Para o cálculo da taxa de analfabetismo funcional, cuja meta consiste em uma redução em 50% até 2024, adotou-se a Pesquisa Nacional por Amostra de Domicílios (Pnad) anual de 2012 como referência. Como a taxa de analfabetismo funcional nesse ano foi de 18,3%, atingir a meta do PNE significa reduzir essa taxa para 9,2% até 2024 (Brasil. Inep, 2016).(Inep,2018).

Para tornar possível e viável a avaliação textual dentro dos níveis de letramento, foi necessário lançar mão da informática como ferramenta capaz de executar, através da programação do software Spekuloom, as funcionalidades de agrupar classes gramaticais em sequência, descobrindo se há sequências que se repetem e, através do *machine learning* ensinar para o computador quais sequências pertencem a cada nível de letramento. O uso da informática é imprescindível para que a análise das classes gramaticais em sequências seja factível.

Por último, mas não menos importante, está a contribuição social que envolve o desenvolvimento do letramento como abordagem que aumenta o nível de pertença em diferentes categorias sociais, já que um indivíduo letrado é capaz de se expressar com coesão e coerência nas diversas situações e âmbitos sociais.

3.2 CRIAÇÃO DO SOFTWARE SPEKULOOM

Após a idealização da proposta do modelo, essa foi colocada em prática através do uso da ferramenta computacional Spekuloom, que é um software livre programado em linguagem Python e salvo no Github com o intuito de ficar a disposição do usuário que tiver interesse de conhecer o código e até o reproduzir se assim for de seu interesse.

Originalmente, o Spekuloom foi estruturado com a intenção do aluno realizar a análise gramatical das palavras de um texto de modo que fosse possível através da interface *drag-and-drop* dos Símbolos Gramaticais Montessori, habilitando a funcionalidade de arrastar e soltar os símbolos abaixo de cada palavra do texto. Porém, ao longo de seu planejamento observou-se que era possível utilizar as sequências de classes gramaticais para avaliar o mesmo texto, logo a ideia originária foi colocada em segundo plano para que a avaliação ficasse em primeiro. Esta análise torna observável qual o nível do texto: básico, intermediário ou transitório; sendo assim uma ferramenta que avalia o nível de letramento em que o escritor se encontra em relação à narrativa.

A elaboração do Spekuloom foi dividida em dez etapas:

- Concepção da ideia de criar um modelo de tecnologia que converge com o ensino Montessori da língua e que usa as novas tecnologias para contribuir com aperfeiçoamento do letramento da escrita de uma narrativa através da avaliação. Um ponto importante é que não reflete uma mera representação digital do que já existe no plano físico como mostra a figura 7, mas um aprofundamento apenas viável através da computação.

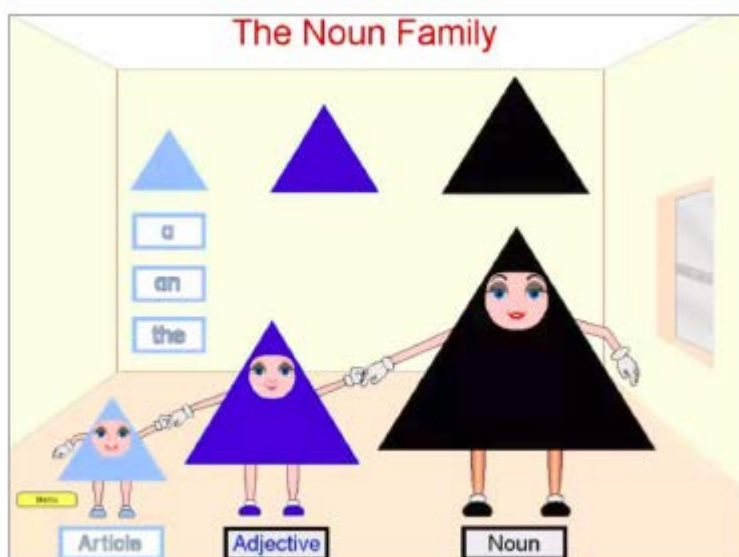


Figura 7. Software Montessori que Transforma Material Concreto em Digital

- Escolha da linguagem Python para a programação da ferramenta computacional Spekuloom pelo uso de sua biblioteca de processamento de linguagem natural *Natural Language Toolkit* e sua propriedade de tokenização exemplificada pela figura 8;

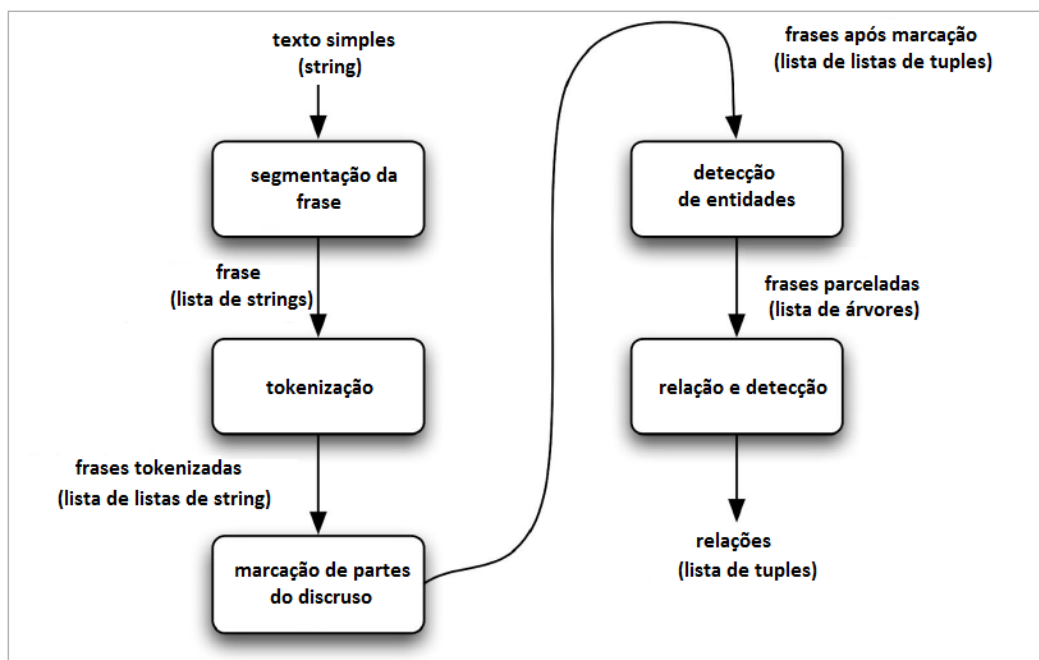


Figura 8. Processo de Tokenização do NLTK

- Divisão em classes gramaticais, como próxima etapa, para que, a cada uma delas, fosse acoplado um símbolo gramatical Montessori. Para as classes que eram muito específicas, novos símbolos foram acrescentados na sua classificação. Nessa etapa, a ferramenta já estava sendo desenvolvida em português;
- Categorização do texto em Símbolos Gramaticais Montessori, possível devido à tokenização feita pelo processamento de linguagem natural. O Spekuloom já era capaz de fazer o papel do aluno, ou seja, realizar a análise gramatical de um texto;

```

Entretanto vou escrever a meus amigos ,
Os cinco atos são uma série de acontecimentos terríveis ,
de atribuições amargas por que a pobre menina teria de pa
Primeiro a desonra ,
Mais tarde quase uma maldição ;
estes sucumbem ,
aqueles suicidam-se ;
é uma procissão de terrores que tem a infelicidade de não
No meio disto tudo ,
dois meliantes que vão à cata de fortuna = e posição ,
que procuram pelo jogo = e pelo assassinato ,
o punhal = e o baralho ,
a cuja invenção deu causa um rei maluco ,
como a bela leitora sabe .
Esses dois varões sem probidade são Picheric = e Roberto ;
Warner = e Julio .
O epílogo começa pela derradeira situação do prólogo ;
o estrangeiro lacra a sua última carta ,
defronte de Henriqueta que se debate em um pesadelo ,
o final do ato 5 . " - Ele levanta-se = e acorda-a .
É uma bela cena .
Henriqueta reconhece a realidade ,
que seus pais estão vivos ,
= e livre de seu sonho terrível abraça-os .

```

Figura 9. Análise Gramatical Simbólica Feita pela Máquina

- Separação de sequências de três a cinco classes gramaticais e observação de quais predominavam em cada gênero textual utilizado para o teste. Escolha da sequência de quatro palavras como principal devido a melhor leitura de resultados mesmo que aumentando uma palavra do padrão original sugerido por Montessori nas famílias.

```

sp : ▲▼▲▲ 19
sp : ■●▲■ 19
sp : ●■▲▲ 19
sp : ■▲▲▲ 19
sp : ▲▲▲▲ 18
sp : ▲▲▲. 16
sp : ■▲▲▲ 15
sp : ▲▲▼▲ 15
sp : ▲▲▼▲ 15

```

Figura 10. Exemplos de Padrões e sua Recursividade

- Escolha do escopo para a dissertação, narrativas de diferentes níveis de letramento: crianças até 6 anos, infantil e juvenil.

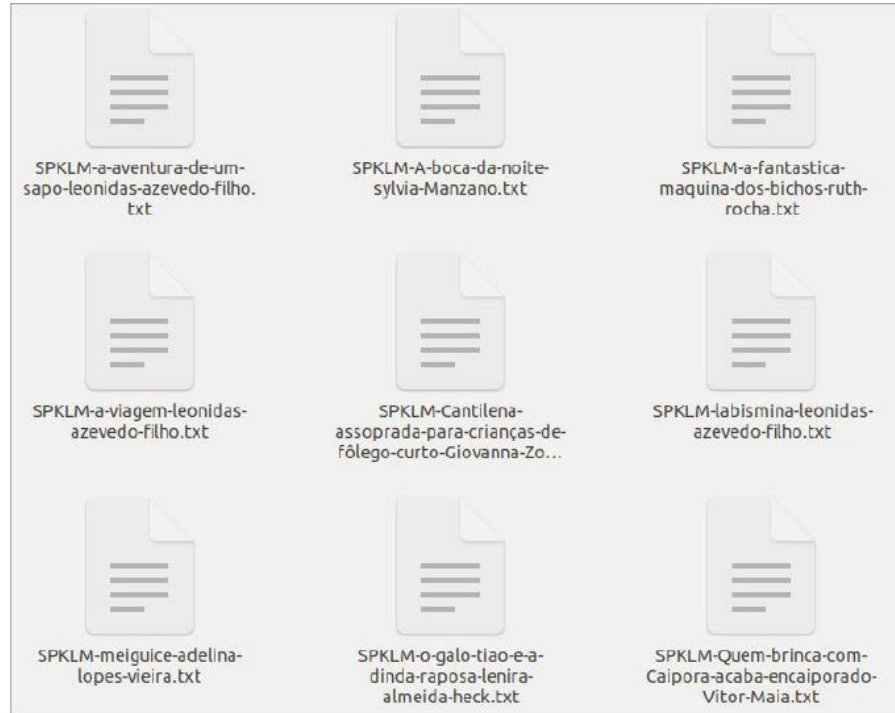


Figura 11. Textos da Categoria Básica

- Aprendizado do sistema a partir dos dados dos livros selecionados, para que através do *machine learning* seja capaz de avaliar textos produzidos pelos alunos com base nos livros utilizados como fonte de dados.

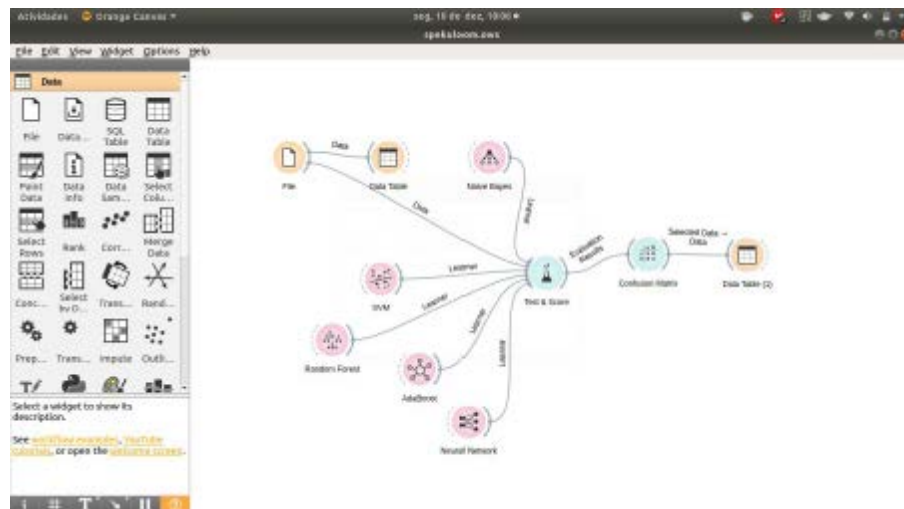


Figura 12. Laboratório Orange

- Utilização de grafos para mostrar *scatter plots* que se formam a partir dos livros analisados. Estes *scatter plots* baseiam-se no cálculo da variância da quantidade de vezes que uma sequência gramatical se repete em cada texto

analisado. Confirmam, assim, a diferença dos textos quanto ao nível de letramento de cada indivíduo.

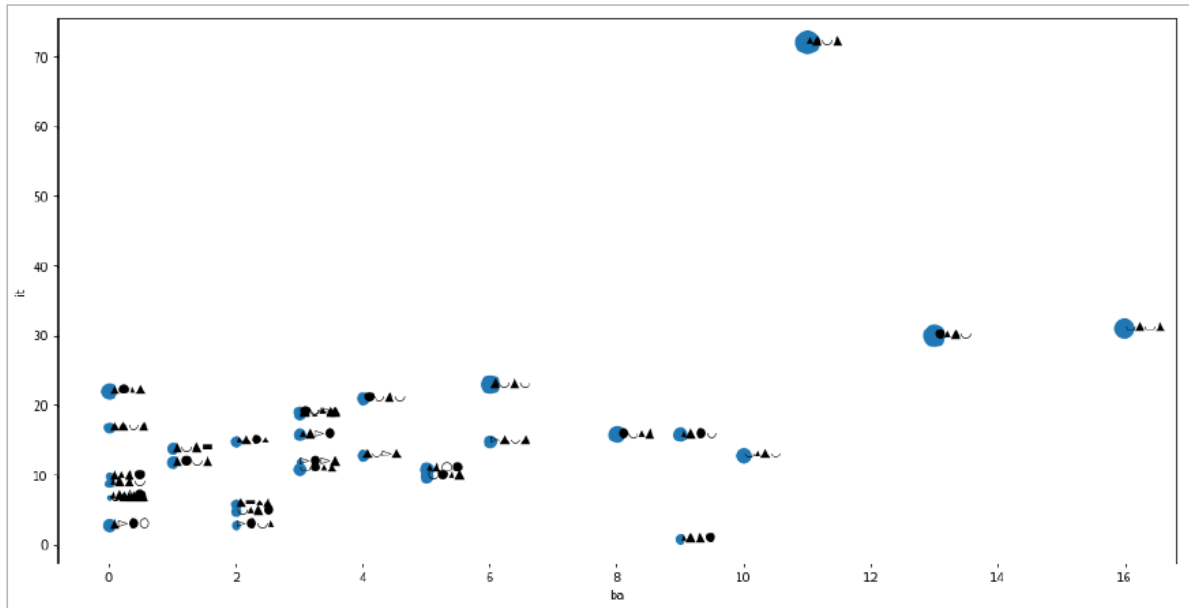


Figura 13. Padrões e sua Recursividade Comparando Níveis Básico e Intermediário

- Análise Bayesiana dos dados de textos narrativos produzidos pelos alunos para avaliação do nível de letramento de cada texto.

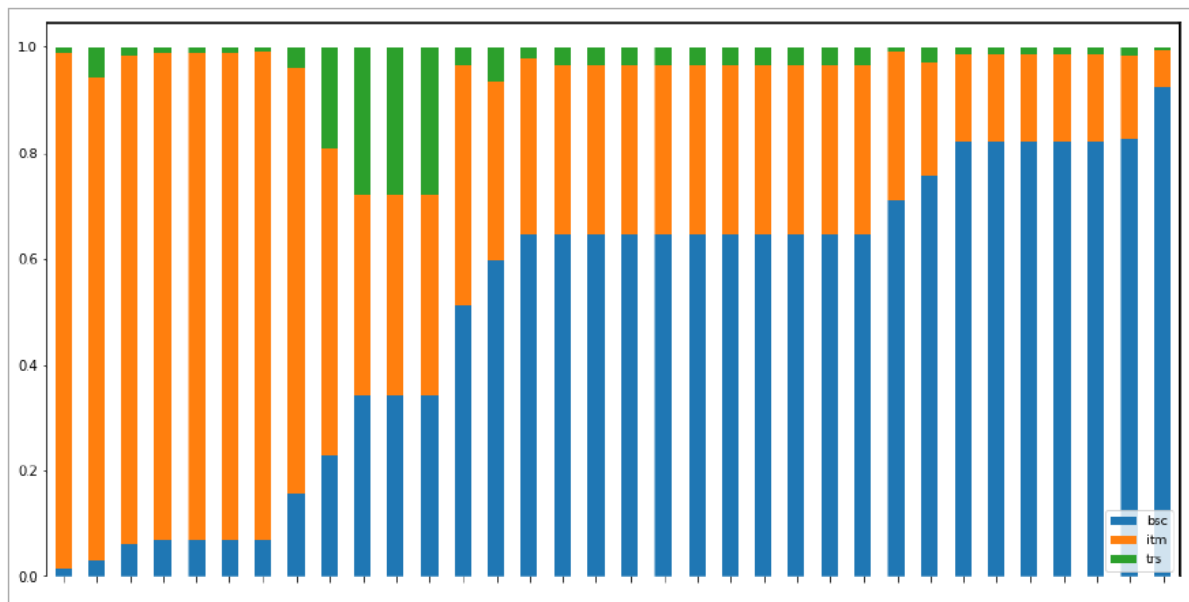


Figura 14. Comparação entre Avaliações Textuais após a Análise Bayesiana

4 METODOLOGIA

O desejo de observar (...) uma harmonia preestabelecida (...) é um estado emocional (...)
semelhante ao de uma pessoa religiosa ou ao de uma pessoa apaixonada.
Albert Einstein

O capítulo descreve o tipo de metodologia utilizado nesta pesquisa como o intuito de discriminar pontos importantes sobre esta tanto do ponto de vista concreto quanto do abstrato.

4.1 TIPO DE PESQUISA

Quanto a seus fins e meios, é uma pesquisa de experimentação, caracterizada por testes com as variáveis em escala reduzida. Prática executada em HTML para a obtenção de dados primários por meio da coleta de informações textuais. O universo desta pesquisa é formado por alunos e professores do ensino fundamental.

4.2 SELEÇÃO

4.2.1 Seleção Textual

Os textos utilizados para ensinar o Spekuloom nessa pesquisa foram retirados do Domínio Público e englobam de cinco a dez textos de cada um dos três níveis de letramento propostos escolhidos:

- Básico: Textos de crianças até 6 anos;
- Transitório: Textos de literatura infantil;
- Intermediário: Textos da literatura juvenil;

4.2.2 Seleção de Indivíduos

Os indivíduos dessa pesquisa são alunos Ensino Fundamental II. Este segmento escolar foi selecionado, pois se trata do período no qual há um olhar mais focado do professor da língua para o desenvolvimento da coesão e coerência do texto.

Para fins deste trabalho, foram utilizados 34 textos de alunos de uma escola Particular de Niterói, Rio de Janeiro, para avaliar a ferramenta Spekuloom quanto à classificação do texto em português em diferentes níveis.

A escolha do local foi determinada pela metodologia Montessori que na escola é aplicada sendo uma facilitadora no entendimento por parte dos professores e alunos ali envolvidos.

4.3 DESCRIÇÃO DO INSTRUMENTO UTILIZADO

Para que houvesse a avaliação textual através da análise da sequência de classes gramaticais, foi necessário criar uma ferramenta computacional denominada Spekuloom. Esta foi programada em linguagem Python, utilizando a biblioteca de processamento de linguagem natural NLTK, tornou possível a tokenização das palavras separar as sequências gramaticais e análise de texto a partir dessas.

4.4 ANÁLISE NO LABORATÓRIO ORANGE

O laboratório Orange foi configurado para treinar máquinas de aprendizado de cinco naturezas diferentes: *Naive Bayes*, *SVM*, *Ada Boost*, *Random Forest* e *Neural Network*.

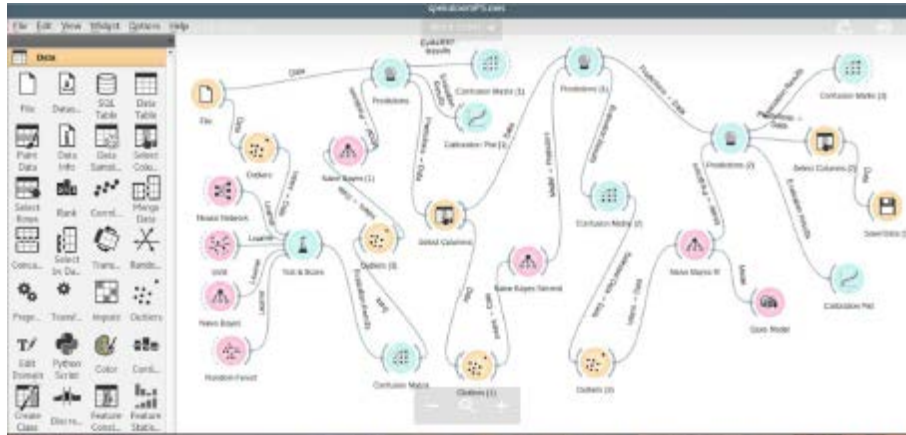


Figura 15. Configuração do Aprendizado de Máquina do Laboratório Orange

O laboratório, também, foi configurado para avaliar a matriz de confusão que relata as diferenças entre os padrões e as classificações obtidas. Na figura 16, podemos ver a matriz de confusão referente ao aprendizado pelo método *Naive Bayes*.

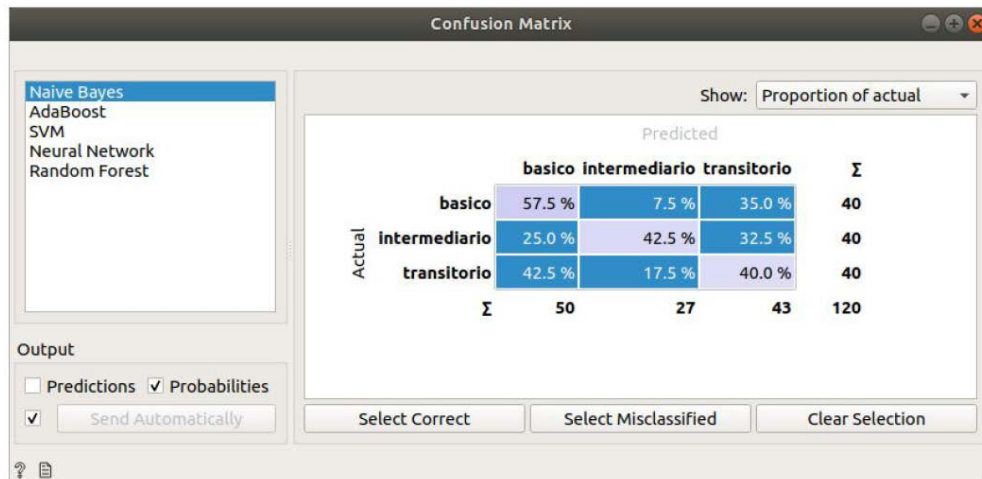


Figura 16. Matriz de Confusão para a Classificação Naive Bayes

5 REALIZAÇÃO DO EXPERIMENTO

Implementação do Spekuloom

A ciência se compõe de erros que, por sua vez, são os passos até a verdade.

Julio Verne

Este capítulo, com base na estratégia de Hockman e Berengut [1995], analisa cada etapa do desenvolvimento de modo a caracterizar a pesquisa experimental, ao mesmo tempo que, explicita e descreve os diferentes estágios do estudo.

5.1 COLETA DE INFORMAÇÕES

Esse estágio caracterizou-se pela busca de informações referentes à classificação textual no que tange a diferenciação entre textos de um mesmo tipo e gênero textual porém com níveis diferentes de elaboração ou seja níveis diferentes de letramento que segundo Soares [2005] é a aptidão de usar o sistema da leitura e escrita para exercer uma prática social em que a sua utilização é necessária. A partir desse pressuposto, a atenção foi voltada para como se classifica o texto no caso da leitura e deparou-se com a Classificação Decimal Universal que serviu de parâmetro para classificar, também, o sistema da escrita.

A primeira etapa de criação da ferramenta tratou da busca por textos modelo que representassem cada nível textual. Estes textos precisavam ser corretos no que diz respeito à ortografia e coerência, porém representariam diferentes etapas da coesão textual.

Aqui o conceito de letramento se aproxima ao entendimento de Kleiman(1995) “Podemos definir hoje o letramento como um conjunto de práticas sociais que usam a leitura e a escrita em eventos relacionados com uso, função e impacto social da escrita”.

Em busca de um critério adequado, e já previamente estabelecido, porém com a necessidade de teoria que a embasasse de modo que não fosse uma escolha arbitrária, encontramos a categorização utilizada em bibliotecas a Classificação Decimal Universal que ordena dentro do universo de livros direcionados ao Ensino Fundamental 1 e Ensino Fundamental 2 - escopo de escritores -desta classificação foram, então, retiradas as seguintes categorias: 82-92 Livros para crianças até aos 6 anos e 82-93 literatura infantil (6 -12 anos) e juvenil (12 -18 anos).

É importante registrar que, após a definição do critério mais adequado de avaliação e a escolha de textos, em sua maior parte, narrativos, partimos para a procura do melhor método de digitalização dos livros. Tivemos o cuidado de selecionar livros de domínio público, testamos digitar, digitalizar, porém, encontrar o melhor meio de obtê-los em formato digital foi encontrá-los em .pdf e transformá-los em .txt. Esse foi o modo menos passível de erros ortográficos e que mais otimizou o tempo.

5.2 DEFINIÇÃO DE OBJETIVOS

Com as classificações do sistema escrito amparado sobre as do sistema da leitura, iniciou-se a etapa de definição de objetivos representados pela busca de como encontrar padrões que caracterizassem cada categoria de classificação em um texto escrito. Neste

estágio, surgiu a *Psicogrammatica* Montessori [2017] que através dos símbolos gramaticais, faz a análise morfológica de um texto.

Em seguida, observou-se que Montessori aglutinava classes gramaticais em padrões de três palavras sequenciais, o que ela denominou família do nome e família do verbo, representados respectivamente pelas figuras 17 e 18.



Figura 17. Família do Nome

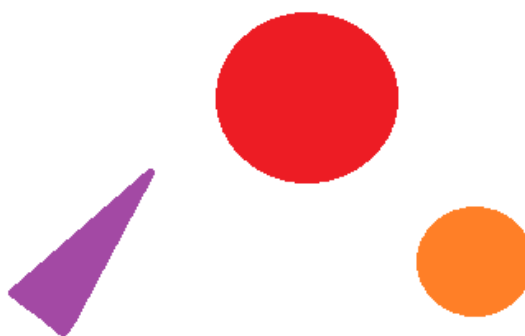


Figura 18. Família do Verbo

Esta investigação, conduziu a pesquisa a seu próximo patamar com o objetivo já traçado, usar aglutinações sequenciais de palavras com base em suas classificações gramaticais para encontrar padrões que configurem cada classificação: Livros para crianças até aos 6 anos, literatura infantil (6 -12 anos) e literatura juvenil (12 -18 anos); que passaram a se chamar, a partir desta etapa, respectivamente: nível básico, nível intermediário e nível transitório.

Para matematizar um texto, no sentido de possibilitar a contagem de padrões e, a partir destes, medir em que nível textual se encontra, foi necessário servir-se da classificação gramatical das palavras, pois esta agrupa palavras de um mesmo tipo, diminuindo, desse modo, as possíveis variáveis.

A gramática na metodologia Montessori é tratada de forma particular. Nela, cada classe gramatical tem um símbolo e cor próprios que são descritos de modo completo em sua *Psicogrammatica*. Além dos símbolos gramaticais, Montessori, também agrupa classes em duas famílias arquetípicas: a família do nome e a família do verbo. Montessori dizia que a

criança, ao analisar um texto, com base nos símbolos gramaticais conseguiria avaliar se este era rico ou não a partir da quantidade de cores. Quanto mais colorido, mais rico ele seria. A visão diferenciada sobre a gramática Montessori foi a grande influenciadora sobre o trabalho aqui realizado.

Com base nas famílias, surgiu-se a ideia da possibilidade de haver outras aglutinações e estas estarem relacionadas a diferentes tipos de texto umas presentes apenas em textos mais elaborados. Assim, surgiram as seguintes perguntas: Há padrões gramaticais em sequência que se repetem em diferentes tipos de texto? Caso haja, é possível utilizar esses padrões para avaliar se um texto é de nível básico, intermediário ou transitório um texto?

5.3 PROJEÇÃO DO EXPERIMENTO

O software Spekuloom foi idealizado nesta etapa como ferramenta que através da linguagem de programação Python, usa o framework NLTK e a Análise Bayesiana para validar a classificação textual, ensinar a máquina a fazer a análise automática e, finalmente, avaliar diferentes textos submetidos ao software. Em um primeiro momento, utiliza o framework NLTK para etiquetar (tokenizar) as classes gramaticais com os símbolos gramaticais Montessori. Em seguida, busca-se por padrões sequenciais de palavras por classes gramaticais em diferentes textos. Ao encontrar os padrões, analisa quais mais se repetem em cada texto. Depois, alimenta os dados com textos das diferentes classificações da CDU usa-se a estatística Bayesiana para prover prognósticos das diversas situações exemplo, ou seja, usa a equação da Análise Bayesiana, apresentada na figura 22, para converter os dados e calcular sobre a recursividade de padrões e sua frequência de probabilidade nos diferentes níveis textuais sendo caracterizado pela a maior porcentagem que qualifica a aprendizagem da máquina. Por fim, textos escritos são submetidos ao programa para serem analisados dentro dos níveis previamente propostos.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Criamos uma ferramenta denominada Spekuloom que viabilizou a pesquisa no intuito de confirmar as hipóteses estabelecidas. Foi construída na linguagem Python por ser uma linguagem aberta e reproduzível. Usamos o framework NLTK, pois nele encontrou-se uma biblioteca apropriada para etiquetar as palavras em símbolos em um primeiro momento apenas em inglês, mas, logo depois, achou-se uma versão simplificada em português que agilizou o trabalho.

Após programar o sistema para etiquetar dentro dos padrões dos símbolos gramaticais Montessori, rodou-se o programa e correções foram necessárias até que o ajuste fosse completo. Em seguida, direcionou-se o programa para separar o texto em diferentes tamanhos de padrões sequenciais, mas, finalmente, concluiu-se que quatro palavras em sequência era o melhor modo de avaliar um texto. Buscou-se, então, algoritmos que investigassem o texto e quantificassem o número de vezes que um agrupamento sequencial aparecia no texto e logo conseguiu-se uma mostra.

A mostra foi o bastante para submeter doze livros das três diferentes categorias para a aprendizagem da máquina. Após a leitura, o programa foi capaz de encontrar alguns padrões - prévia apresentação do que seria capaz de fazer - porém, quando submetidos à Análise Bayesiana, não foi o bastante para aproximar os dados aos cem por cento esperados. A conclusão desta análise levou à busca por novos livros das diferentes categorias até que se encontrou uma análise próxima ao desejado.

Após a calibragem da máquina, foi possível apresentar textos escritos por alunos ao software. Este os comparou aos diferentes padrões encontrados nos livros previamente nivelados e conseguiu avaliar o texto a partir dos padrões que ele apresenta, provando que a hipótese era verdadeira. É possível encontrar padrões gramaticais que se repetem sequencialmente em um texto e utilizá-lo para avaliar a riqueza da escrita de um texto.

5.4 OPERAÇÃO DO EXPERIMENTO

O primeiro experimento após a alimentação de dados para a aprendizagem da máquina com quarenta livros classificados nas diferentes categorias da CDU foi realizado em dez textos que continham entre 700 e 1800 caracteres.

Depois de calibrar a máquina com os livros dos diversos níveis, dez textos de alunos foram selecionados para servirem como grupo piloto. Ao observar os resultados, descobriu-se que surgia, de modo inesperado, um número único de resultados.

Analisar textos de diferentes tamanhos fez com que a quantidade de vezes que cada padrão aparecia nos livros fosse imensamente maior do que as encontradas nos textos escritos pelos alunos, categorizando, desta maneira, quase sua totalidade em textos de níveis básicos.

5.5 ANÁLISE DO EXPERIMENTO

O software Spekuloom foi utilizado para realizar a análise do texto dentro do critério de padrões simbólicos Montessori. Como resultado, apresentou sete diferentes padrões pertinentes a cada um dos dez textos e sua recursão em cada texto entre colchetes representados pela figura 19.

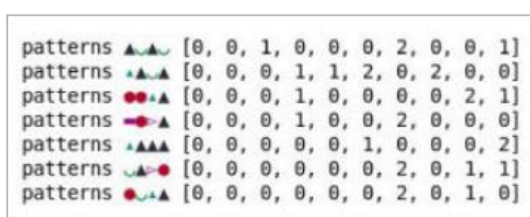


Figura 19. Análise de Padrões Gramaticais de Dez Textos Submetidos

O Spekuloom foi capaz de rodar sua avaliação e permitir que os padrões das classes gramaticais surgissem a partir dos textos. A análise dos resultados apresentou a recursão pelo menos duas vezes de sete diferentes padrões, dentre os dez textos encontrados, um número pequeno se comparado à quantidade de textos.

Os alunos apresentavam textos de três a cinco parágrafos, o que dava para especular a presença de uma quantidade um pouco maior do que a apresentada. Então observamos que haviam poucos textos, apenas dez e, além disso, os livros apresentavam escritas com uma fração muito maior de parágrafos.

Ao comparar os resultados dos alunos - sete padrões - com os dos livros - trinta e dois padrões- sabíamos que era necessário realizar alguma modificação na estrutura de dados para que ficasse favorável para a avaliação correta desses textos, logo resolvemos adaptar os dados.

5.6 INTERPRETAÇÃO DOS RESULTADOS

Ao observar apenas sete padrões encontrados, em comparação à análise dos livros – trinta e dois padrões – logo observou-se a necessidade de incluir novos textos para que novos padrões surgissem e pudessem ser analisados de forma mais apurada.

Foi feita uma reavaliação nos livros já apresentados à máquina, limpando os dados inadequados e reajustando alguns livros em níveis mais apropriados. A máquina foi

alimentada com mais dezenove novos livros para estruturar melhor os padrões encontrados e voltamos a realizar novos testes.

5.7 ADEQUAÇÃO DO EXPERIMENTO

Foi necessário analisar trinta e quatro textos para que a quantidade de padrões encontrados aumentasse para vinte, representado pela figura 20, número significativo e propício para que haja a avaliação piloto de textos com base no aprendizado feito pela máquina.

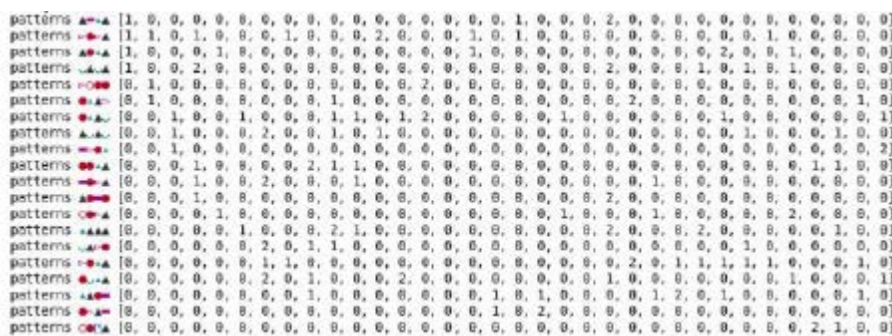


Figura 20. Análise de Padrões Gramaticais de Trinta e Quatro Textos Submetidos

Neste experimento, a máquina foi alimentada, inicialmente, com dez textos, em seguida com vinte textos, depois, com trinta textos até chegar a trinta e quatro textos, ideal para encontrar a quantidade de padrões apropriados para realizar uma comparação entre os textos de alunos dos diferentes níveis. Para chegar a vinte exemplares de padrões, foi necessário triplicar a quantidade de textos submetidos indicando a dificuldade de se deparar com padrões mais refinados.

5.8 OPERAÇÃO DE CONFIRMAÇÃO DO EXPERIMENTO

De posse dos resultados, realizamos a rodada de confirmação para verificar se o modelo estatístico previsto foi confirmado. Ao enviar os trinta e quatro textos, observou-se que todos com a exceção de um, figura 21, eram considerados de nível básico de acordo com a avaliação.

38 instances	11 sample_25.t.	basic	0.906	0.002	0.003	0	0	0	1	0
20 features (no missing values)	12 sample_26.t.	basic	0.926	0.006	0.006	0	0	0	0	0
No target variable	13 sample_27.t.	basic	0.907	0.009	0.004	0	0	0	1	0
5 extra attributes (20.0% missing values)	14 sample_28.t.	basic	0.906	0.002	0.008	0	2	0	2	0
	15 sample_17.t.	basic	0.905	0.002	0.003	0	0	0	0	0
	16 sample_25.t.	basic	0.901	0.002	0.007	0	0	0	0	0
Variables	17 sample_4.t.	basic	0.905	0.002	0.003	0	0	0	0	0
	18 sample_24.t.	basic	0.933	0.007	0.009	0	0	0	0	0
<input checked="" type="checkbox"/> Show variable labels (if present)	19 sample_13.t.	basic	0.905	0.002	0.003	0	0	0	0	0
<input checked="" type="checkbox"/> Missing numeric values	20 sample_15.t.	basic	0.905	0.002	0.003	0	0	0	1	0
<input checked="" type="checkbox"/> Color by instance classes	21 sample_40.t.	basic	0.905	0.002	0.003	0	0	0	0	0
	22 sample_41.t.	basic	0.817	0.033	0.230	2	0	0	0	0
	23 sample_51.t.	basic	0.905	0.002	0.003	0	0	2	0	0
Selective	24 sample_23.t.	basic	0.906	0.002	0.003	0	0	0	0	0
	25 sample_18.t.	basic	0.906	0.002	0.003	0	0	0	0	0
<input checked="" type="checkbox"/> Select Follows	26 sample_44.t.	basic	0.907	0.002	0.002	1	0	0	0	0
	27 sample_6.t.	basic	0.904	0.006	0.002	0	0	0	1	0
	28 sample_14.t.	basic	0.900	0.008	0.002	1	0	0	0	0
	29 sample_12.t.	basic	0.902	0.002	0.006	0	0	0	0	0
	30 sample_7.t.	basic	0.908	0.008	0.004	1	0	0	0	0
	31 sample_28.t.	basic	0.906	0.002	0.003	0	0	0	0	0
Restore Original Order	32 sample_20.t.	basic	0.929	0.008	0.003	0	0	0	0	0
	33 sample_29.t.	basic	0.906	0.002	0.003	0	0	1	0	0
	34 sample_1.t.	basic	0.926	0.002	0.004	0	0	0	1	2

Figura 21. Avaliação Textual Subdividido em Básico, Intermediário e Transitório

Os caminhos indicam que havia algum erro na classificação textual, provavelmente devido à quantidade de caracteres utilizados para a busca de padrões nos livros, aproximadamente 8000, em comparação a quantidade de caracteres utilizados na busca de padrões nos textos, de 700 – 1800.

Assim, submeteu-se um maior número de livros e para aumentar sua precisão e fez-se um recorte, pois observou-se que os livros estavam oferecendo uma maior gama de padrões e de repercussões e no momento de avaliar os textos curtos não se encontrava tamanha repercussão ou diversidade, trazendo sua avaliação textual mais para próxima do básico. Decidiu-se diminuir a quantidade de caracteres utilizados nos livros para um número próximo ao apresentado pelos textos dos alunos. Essa modificação, do tamanho do texto, acabou interferindo na Análise Bayesiana, aproximando-se cada vez mais dos cem por cento esperados.

5.9 APLICAÇÃO DOS RESULTADOS

Mesmo a análise de resultados ainda não completamente calibrada, é possível aplicar o resultado nos textos, pois ao avaliar o exemplo de nível transitório, este apresenta uma construção mais elaborada do uso da língua portuguesa em relação aos outros textos, indicando que a hipótese de encontrar padrões recorrentes e pertinentes a cada nível textual, está sendo comprovada aproximando-se assim do objetivo do estudo de usar esses padrões para classificar os textos.

Observe o contraste entre os trechos básico e transitório respectivamente a seguir:

Na minha opinião morar nas grandes cidades, eu acho não tem muitos problemas sociais, econômicos ou civis, mas isso não significam que não exita crimes nas grandes

ciudades, em todo lugar existe crime, mesmo que aparente não ter, uma vez inclusive quando estava voltando do trabalho,[...]

Enxergar as grandes cidades como florestas, onde as pessoas são animais perdidos e abatidos e os prédios são grandes e misteriosas árvores. A indiferença e o medo estampam os rostos pela rua e a necessidade e os excessos se contrastam.

6 ANÁLISE DE RESULTADOS

Se quisermos alcançar resultados nunca antes alcançados, devemos empregar métodos antes nunca testados.

Francis Bacon

No atual capítulo, se fará uma análise de cada resultado apresentado, sobre o modelo de avaliação textual oportunizado pelo software Spekuloom.

Criar uma fermenta computacional que permitisse um novo parâmetro de avaliação textual envolveu diferentes etapas de elaboração. Entre elas, a transformação computacional de palavras em símbolos gramaticais Montessori, etapa que iniciou toda pesquisa. Passada essa etapa da programação, foi necessário agrupar palavras em diferentes quantidades até que se optou por um grupo de quatro, devido a sua melhor visualização quando há necessidade da observação do fenômeno a olho nu.

Com o parâmetro já pré-estabelecido, voltou-se a atenção para o ensino da máquina. Foi necessário pesquisar por um padrão de separação textual em vigor para caracterizar os diferentes níveis de letramento da escrita de narrativas e, para isso, seguimos a categorização utilizada em bibliotecas, a Classificação Decimal Universal (CDU), e separamos em três categorias adequadas para a idade em questão, de seis à quatorze anos, respectivamente básico, intermediário e transitório. A partir daí, foram angariados livros de domínio público no formato .pdf para servirem de exemplos de padrões encontrados em cada categoria.

Por fim, foram submetidos trinta e quatro textos de alunos do Ensino Fundamental 2 para a avaliação feita pelo Spekuloom já, inicialmente, calibrado pelos textos de domínio público. A análise dos resultados se manterá sobre as avaliações textuais realizadas pela ferramenta computacional, indicando sua viabilidade nesta tarefa.

6.1 AVALIAÇÃO TEXTUAL E RESULTADOS

A primeira avaliação trata-se do teste feito para observar se era possível encontrar padrões que se repetissem em diferentes tipos textuais escolhidos de forma aleatória para fins de teste do espelhamento simbólico em padrões recursivos foram submetidos ao Spekuloom que analisou e plotou um gráfico com a quantidade de vezes que os padrões sequenciais de classes gramaticais mais recorrentes se apresentaram. Para o gráfico da figura 22, foram utilizados padrões de quatro símbolos gramaticais em sucessão.

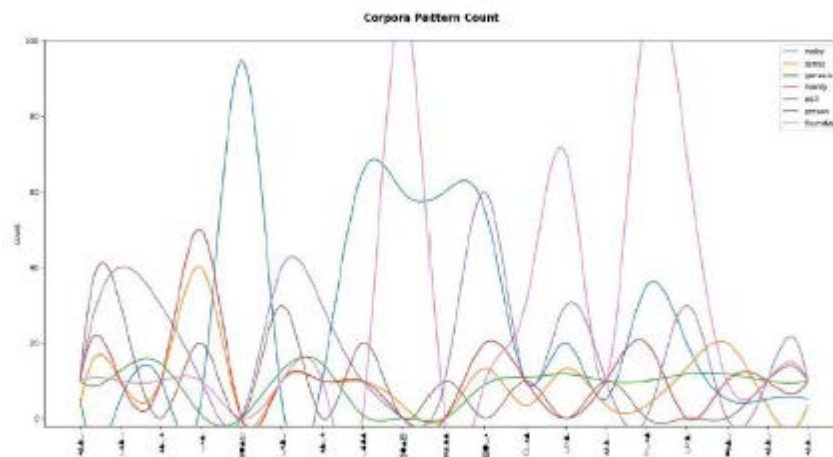


Figura 22. Padrões Recursivos em Grupos de Quatro Palavras

É importante observar que entre o marco zero e vinte, no eixo de contagem do gráfico, existe uma recorrência de padrões mesmo que em obras de gêneros textuais distintos. Outro ponto crucial é observar que em certos textos, simbolizados pelas linhas coloridas, padrões, símbolos em grupos de quatro na base do gráfico, se repetem no próprio escrito de forma numerosa podendo sinalizar um padrão inerente a este tipo de escrita.

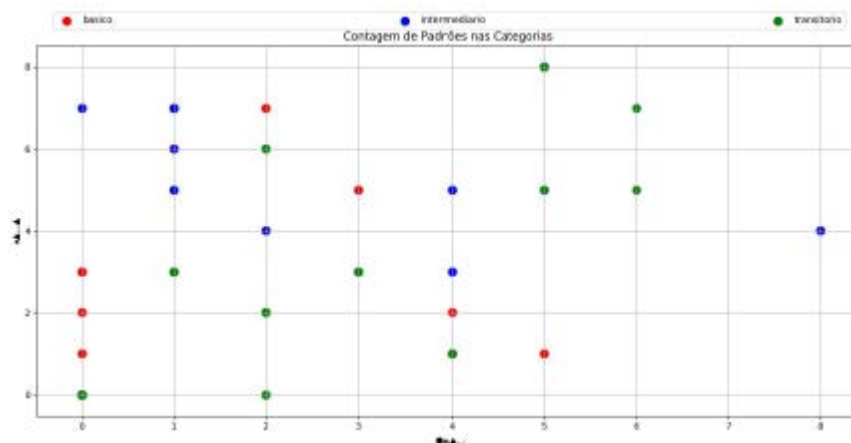


Figura 23. Scatter Plot de Padrões em Diferentes Categorias

Na figura 23, podem-se observar agrupamentos de diferentes textos analisados quando sob a perspectiva de duas sequências de classes gramaticais, ou seja, a recorrência de cada padrão indica a qual nível de letramento (básico, intermediário ou transitório) o texto pertence.

A implicação pedagógica principal desse achado é que se há uma recorrência de padrões de classes gramaticais sequenciais em diferentes textos é possível utilizá-los para viabilizar uma estratégia observável de letramento.

A identificação dos padrões de contagem associados aos gêneros textuais poderá ser usada, em trabalhos futuros, para uma ferramenta de tutoria automática. Textos de diversos

níveis de letramento poderão ser fornecidos a um processo de aprendizagem de máquina que poderá monitorar avanço do estudante e fornecer *scaffolding* para o aprendizado da língua.

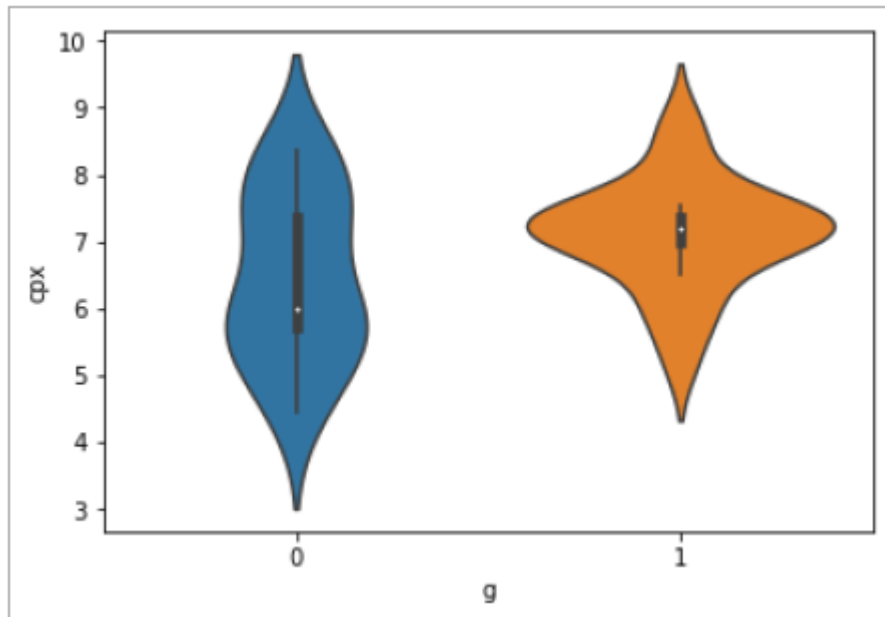


Figura 24. *Scatter Plot* da Avaliação Textual dos Alunos separados por Gênero: Feminino e Masculino
 Quanto à avaliação dos textos dos alunos, observa-se que, com base nos padrões, há uma distribuição marcante entre a pontuação das narrativas dentro do próprio gênero, enquanto que no gênero masculino, grande parte se concentra na pontuação sete. O Spekuloom é passível de fazer diferentes análise do estilo, podendo servir como base para diferentes estudos ligados à educação em trabalhos futuros.

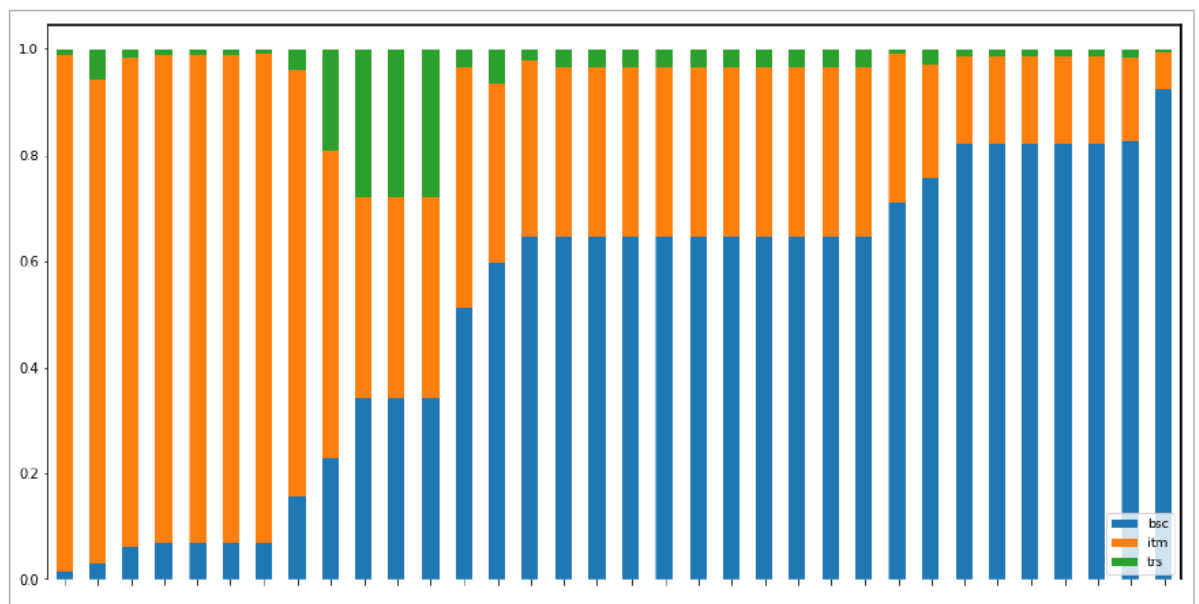


Figura 25. *Scatter Plot* da Quantidade de Padrões de Diferentes Níveis Presente no Texto de Cada Aluno

Na figura 25, cada linha vertical se refere a um aluno e cada cor a um nível de letramento da escrita da narrativa. Cor azul representa o nível básico, a cor amarela representa o nível intermediário e a cor verde representa o nível transitório, ou seja, cada linha vertical representa a quantidade de padrões que se repetem característicos de cada nível, partindo dessa base para criar a avaliação de cada aluno. Esta imagem representa o resultado final de todo o estudo, confirmando a possibilidade de haver uma avaliação com base nos critérios adotados nesta pesquisa.

7 CONCLUSÃO

Quanto aos homens, não é o que eles são que me interessa, mas o que eles podem se tornar.

Jean Paul Sartre

A conclusão retoma os pontos principais de cada capítulo, finalizando com a importância da contribuição da pesquisa realizada no presente estudo.

Essa dissertação de mestrado assumiu como objetivo criar uma ferramenta computacional que servisse de apoio à avaliação de redações feitas por professores, baseando-se em padrões gramaticais recursivos em diferentes níveis de letramento da escrita de narrativas. Para que os resultados fossem de melhor compreensão aos olhos humanos, foram utilizados os símbolos gramaticais Montessori para realizar a marcação das classes gramaticais.

Para tal, esta análise apoiou-se em três pilares fundamentais para seu processo de desenvolvimento científico: informática, educação e sociedade. O primeiro, responsável pelo papel da máquina de otimizar tarefas que demorariam um tempo infinitamente maior se feito pelas mãos humanas. O segundo, de onde surgiu o problema, marca a necessidade de uma concretude maior na avaliação de textos escritos para que seja possível visualizar uma escala evolutiva destas através da metalinguagem. O terceiro, de modo indireto, permite que o professor tenha uma ferramenta que o apoie em sua correção, de modo que a partir desta consiga expressar com maior rigor quais pontos de coesão textual podem ser aprimorados pelos alunos para que, assim, avancem em seu nível de letramento da escrita.

Realizou-se, em primeiro lugar, uma revisão sistemática sobre a educação e a tecnologia. Apesar do uso da tecnologia na educação ser estudado há algum tempo, esta pode ser subutilizada em sala de aula principalmente quando se trata da internet. As opiniões quanto aos benefícios de seu uso são divididas, mas é visível que faz o papel de mediadora do processo de comunicação, pois oferece vantagens para a organização, compartilhamento e registro de informações, além da oferta de ferramentas para a comunicação e construção de textos coletivos. Há também o diferencial que marcou no ensino a distância, agilizando e qualificando os processos de aprendizado.

O experimento baseou-se no modelo de espelhamento das palavras em suas classes gramaticais, no agrupamento destas classes em padrões recursivos de quatro palavras e na marcação da frequência desses padrões em diferentes níveis de letramento de escrita narrativa, sendo esta frequência a responsável por avaliar o nível do texto do aluno. Todo este modelo só foi possível a partir da criação da ferramenta Spekuloom que viabilizou, através da programação em Python com o uso da plataforma NLTK, este processo.

Concluiu-se, ainda, que a presente pesquisa é um exemplo de uso da tecnologia de modo a aprofundar os estudos sobre educação no que diz respeito ao letramento da escrita com o escopo da narrativa apenas possível através do viés tecnológico.

Ainda que o estudo esteja concentrado apenas em um escopo pequeno dentro de todas as possibilidades de escrita, ele pode se tornar um modelo para a reprodução em outros tipos e gêneros textuais.

Para finalizar, este estudo não substitui a tarefa do professor de correção de redações, mas o apoia matematicamente no que diz respeito à avaliação da coesão textual, localizando o autor do texto em um dos três níveis de competência da escrita, para que o professor, a partir daí, possa criar estratégias para que o aluno consiga evoluir a sua escrita.

O presente estudo apresenta resultados concretos de que é possível utilizar a metalinguagem como caminho lógico-matemático para medir o nível de coesão de um texto, contribuindo assim, de forma indireta, com a aquisição do letramento da escrita.

Se o Spekuloom for utilizado pelos professores, eles poderão tê-lo como referência quando houver alguma dúvida quanto à qualidade de sua correção, tirando esta do campo apenas subjetivo e a trazendo para a qualidade métrica.

Apresenta algumas limitações, nomeadamente à quantidade de textos submetidos, demarcando uma pequena amostra do que é possível ser coletado. O software Spekuloom ainda não está pronto para ser utilizado pelo público leigo, o que caracteriza outra limitação. Por fim, a pesquisa aborda apenas a avaliação textual, sendo tarefa do professor sugerir novas estratégias para que o aluno se desenvolva caracterizando uma função indireta no desenvolvimento do letramento.

Apesar das limitações identificadas, e de outras que poderão ser apontadas, considera-se que o estudo permitiu um novo olhar para critérios de avaliação de textos escritos voltados para a coesão textual.

Futuras investigações poderiam sugerir, ao autor, padrões gramaticais recursivos provenientes do nível seguinte para que na sua próxima escrita tentasse se utilizar destes para progredir na aquisição do letramento da escrita de narrativas.

Sugere-se, também, um alargamento do escopo escolhendo novos tipos textuais e novos gêneros, para observar se os mesmos padrões se repetem independentes destas categorias ou não.

Por fim, este estudo constituiu apenas um contributo para que a avaliação textual fosse realizada de forma mais natural, ou seja, menos massacrante para os profissionais da área do

mesmo modo que oferecesse uma nova e possível visão sobre a coesão textual, para que esta seja estudada posteriormente com mais aprofundamento.

Referências

A Redação no Enem 2019 – *Cartilha do Participante*. INEP, 2019. Disponível em <http://inep.gov.br/informacao-da-publicacao/-/asset_publisher/6JYIsGMAMkW1/document/id/6736715>

ALMEIDA, A. C.; SANTOS, M. eds. CDU: Classificação Decimal Universal: tabela de autoridade. Biblioteca Nacional Portuga, 2005.

BARROS, D. R. et al. O Jogo do Jornal: construindo novas estratégias de Letramento. In: SBIE, 20. Florianópolis, 2019

BERTONCELLO, L.; BIANCONCINI, M. E. A. Ensinando com tecnologia no passado e no presente: dois momentos do projeto Apple Classrooms of Tomorrow (ACOT). *Ciências & Cognição*, v. 15, n. 1, p. 33-42, 2010.

BIRD, S.; LOPER, E.; KLEIN, E. *Natural Language Processing with Python*. O'Reilly Media Inc, 2009.

CASAÑAS, J. T.; BLANCH, R. C. Métodos para medir la riqueza léxica de los textos. *Revisión y propuesta*. *Verba: Anuario Galego de Filoloxía* 44: 347-408, 2017.

FAGUNDES, L. *Informática e educação*, Rio de Janeiro: UFRJ/NCE, 1988.

FERNANDES, A. M. *Brasil no PISA 2015: Análises e reflexões sobre o desempenho do estudantes brasileiros*. Fundação Santillan, 2016.

GOODWIN, L. D.; GOODWIN, W. L.; GAREL, M. B. Use of microcomputers with preschoolers: A review of the literature. *Early Childhood Research Quarterly*, v. 1, n. 3, p. 269-286, 1986.

HOCKMAN, K.K.; BERENGUT, D. Design for experiments. *Chemical engineering*, November, pages 142-147, 1995.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). *Relatório do 2º ciclo de monitorato das metas do Plano Nacional de Educação – 2018*. Brasília: MEC/Inep, 2018.

JOLY, M. C. R. A. *Tecnologia No Ensino: Implicações Para A Aprendizagem*. Casa do Psicólogo, 2002.

KLEIMAN, A. Modelos de letramento e as práticas de alfabetização na escola. In: ___. (Org.). *Os significados do letramento: uma nova perspectiva sobre a prática social da escrita*. Campinas: Mercado de Letras, 1995.

KNOX, J. *Archetype, attachment, analysis: Jungian psychology and the emergent mind.* London and New York: Routledge, 2005

MATEUS, M. H. M. *Gramática da língua portuguesa: elementos para a descrição da estrutura, funcionamento e uso do português actual*. Livraria Almedina, 1983.

- MONTESSORI, M. Dr. Montessori's own handbook. Schocken, 2011.
- MONTESSORI, M. *Mente absorvente*. Nordica, 1987.
- MONTESSORI, M. *Psicogrammatica*. Franco Angeli, 2017.
- MONTESSORI, M. *Spontaneous activity in education*. Montessori Helper, 2014.
- MONTGOMERY, D.C. *Design and Analysis of Experiments* 3rd Ed. New York: Jhon Wiley and Sons, 1991.
- PIAGET, J.; CABRAL, A.; OITICICA, C. M. *A formação do símbolo na criança: imitação, jogo e sonho, imagem e representação*. Terceira Edição, 1964.
- PIAGET, J. *O nascimento da inteligência na criança*. Trad. Álvaro Cabral. Rio de Janeiro: Zah, 1975.
- RAMOS, D. K. *Sobre professores, colaboração e tecnologias: reflexões sobre os processos colaborativos e o uso da tecnologia na educação*. *Educação Temática Digital*, v. 9, n. 1, p. 375, 2007.
- SASI, S. *Creating Algorithmic Symbols to Enhance Learning English Grammar*. *International Journal of Research in English Education*, 3 (2), 69-93, 2018.
- SAUSSURE, F. *Premier Cours de Linguistique Générale (1907): d'après les cahiers d'Emile Constantin / Saussure's third course of lectures on general linguistics (1907): from the notebooks of Emile Constantin*. French text edited by Eisuke Komatsu e English text edited by Roy Harris. Inglaterra: Pergamon Press, 1996 [1907]. p. 166.
- SOARES, M. B.; BATISTA, A. A. G. *Alfabetização e letramento: Caderno do Professor*. Belo Horizonte: Ceale/FaE/UFGM, 64 p. (Coleção Alfabetização e Letramento), 2005.
- SILVEIRA, S. A. *Exclusão Digital: a miséria na era da informação*. São Paulo: Fundação Perseu Abramo, 2001.
- SOARES, M. B. *Letramento. Um tema em três gêneros*. Belo Horizonte: Autêntica, 2009.
- SOUSA, R. P. de et al. *Teorias e práticas em tecnologias educacionais*. 2016.
- SOUZA, R. R. *Algumas considerações sobre as abordagens construtivistas para a utilização de tecnologias na educação*. *Liinc em revista*, v. 2, n. 1, 2006.
- TFOUNI, L. V. *Letramento e alfabetização*, 2002.
- VAL, M.G.C. *Redação e Textualidade*. São Paulo: Martins Fontes, 1994.
- VALENTE, J. A. *Por que o computador na educação?* In: SALGADO, M. U. C. *Tecnologias da educação: ensinando e aprendendo com as TIC*. Brasília: MEC/SEED, 2008.
- WAGNER, W. et al. *Natural language processing with python, analyzing text with the natural language toolkit*. *Language Resources and Evaluation*, v. 44, n. 4, (p. 421-424), 2010.

Anexos

ANEXO A – DESCRIÇÃO E PERCENTUAL DE ESTUDANTES NOS SETE NÍVEIS DE PROFICIÊNCIA EM LEITURA – PISA 2015

Nível	Escore mínimo	Percentual de estudantes no nível	Características das tarefas
6	698	OCDE: 1,11% Brasil: 0,14%	Nesse nível, as tarefas normalmente requerem que o leitor faça múltiplas inferências, comparações e contrastes com precisão e detalhamento, e que demonstre a compreensão completa e detalhada de um ou mais textos, podendo envolver a integração de informações de mais de um texto. Podem exigir que o leitor lide com ideias desconhecidas, na presença de informações concorrentes relevantes, e produza categorias abstratas para interpretação. Tarefas de refletir e analisar podem solicitar que o leitor levante hipóteses sobre ou avalie criticamente um texto complexo sobre um assunto desconhecido, levando em consideração critérios ou perspectivas múltiplos e aplicando interpretações sofisticadas externas ao texto. Uma condição marcante para tarefas de localizar e recuperar nesse nível é a precisão da análise e a atenção refinada a detalhes pouco perceptíveis nos textos.
5	626	OCDE: 7,22% Brasil: 1,31%	Nesse nível, tarefas de recuperação de informação requerem que o leitor localize e organize informações profundamente integradas, inferindo sobre quais informações no texto são relevantes. Tarefas de refletir pedem avaliação crítica ou levantamento de hipóteses, com base em conhecimento especializado. Tanto tarefas interpretativas como reflexivas exigem uma compreensão total e detalhada de texto com conteúdo ou forma não familiar. Para todos os aspectos da leitura, as tarefas nesse nível normalmente envolvem lidar com conceitos contrários às expectativas.
4	553	OCDE: 20,45% Brasil: 6,36%	Nesse nível, tarefas de recuperação de informação requerem que o leitor localize e organize diversos fragmentos de informação integrada. Algumas tarefas nesse nível exigem interpretação do significado de nuances da linguagem em uma parte do texto, levando em consideração o texto como um todo. Outras tarefas interpretativas nesse nível exigem que o leitor use conhecimento público ou formal para levantar hipóteses ou analisar criticamente um texto. O leitor deve demonstrar uma compreensão precisa de textos longos ou complexos cujo conteúdo ou forma podem não ser conhecidos.
3	480	OCDE: 27,91% Brasil: 16,19%	Nesse nível, as tarefas requerem que o leitor localize e, em alguns casos, reconheça a relação entre vários fragmentos de informação que devem satisfazer múltiplas condições. Tarefas interpretativas exigem que o leitor integre várias partes do texto a fim de identificar a ideia principal, entender a relação ou construir o significado de uma palavra ou oração. O leitor deve considerar muitas características textuais ao fazer comparações, diferenciações e categorizações. Em geral, a informação exigida não é relevante, há muita informação concorrente ou o texto apresenta outros obstáculos, tais como ideias contrárias à expectativa ou formuladas de maneira negativa. Tarefas reflexivas nesse nível podem solicitar correlações, comparações e explicações ou exigir que o leitor avalie uma característica do texto. Algumas exigem que o leitor demonstre uma compreensão refinada do texto em relação a conhecimentos do cotidiano. Outras tarefas não requerem uma compreensão detalhada do texto, mas pedem que o leitor explore um conhecimento menos comum.
2	407	OCDE: 23,24% Brasil: 25,00%	Nesse nível, algumas tarefas requerem que o leitor localize um ou mais fragmentos de informação, que podem ter de ser inferidos ou satisfazer diversas condições. Outras exigem o reconhecimento da ideia principal em um texto, o entendimento de relações ou a construção de significado dentro de uma parte específica dele quando a informação não é proeminente e o leitor deve fazer inferências de nível baixo. Tarefas nesse nível podem envolver comparação ou contraste com base em uma característica única do texto. Tarefas típicas de reflexão exigem que o leitor faça uma comparação ou diversas correlações entre o texto e o conhecimento externo, explorando sua experiência e atitudes pessoais.
1a	335	OCDE: 13,59% Brasil: 26,52%	Nesse nível, as tarefas requerem que o leitor localize um ou mais fragmentos independentes com informação explícita, reconheça o assunto principal ou a finalidade do autor em um texto sobre assuntos conhecidos ou faça uma correlação simples entre a informação no texto e um conhecimento do cotidiano. Normalmente, a informação exigida no texto é evidente e há pouca, ou nenhuma, informação concorrente. O leitor é explicitamente direcionado a considerar os fatores relevantes na tarefa e no texto.
1b	262	OCDE: 5,23% Brasil: 17,41%	Nesse nível, as tarefas requerem que o leitor localize um único fragmento de informação explícita em uma posição evidente em um texto curto e sintaticamente simples, com contexto e tipo de texto conhecidos – por exemplo, uma narrativa ou uma lista simples. O texto normalmente fornece ajuda para o leitor, tal como a repetição da informação, apresentação de figuras ou símbolos conhecidos. Há um mínimo de informação concorrente. Em tarefas que exigem interpretação, o leitor pode precisar fazer correlações simples entre fragmentos de informações adjacentes.
Abaixo de 1b		OCDE: 1,25% Brasil: 7,06%	A OCDE não especifica as habilidades desenvolvidas.

Fonte: OCDE, INEP.

Tabela 1 . PISA 2015 – Parâmetro para Avaliar Proficiência em Leitura

ANEXO B – AVALIAÇÃO DA REDAÇÃO DO ENEM – COMPETÊNCIA

1

COMPETÊNCIA 1

DEMONSTRAR DOMÍNIO DA MODALIDADE ESCRITA FORMAL DA LÍNGUA PORTUGUESA

200 pontos	Demonstra excelente domínio da modalidade escrita formal da língua portuguesa e de escolha de registro. Desvios gramaticais ou de convenções da escrita serão aceitos somente como excepcionalidade e quando não caracterizarem reincidência.
160 pontos	Demonstra bom domínio da modalidade escrita formal da língua portuguesa e de escolha de registro, com poucos desvios gramaticais e de convenções da escrita.
120 pontos	Demonstra domínio mediano da modalidade escrita formal da língua portuguesa e de escolha de registro, com alguns desvios gramaticais e de convenções da escrita.
80 pontos	Demonstra domínio insuficiente da modalidade escrita formal da língua portuguesa, com muitos desvios gramaticais, de escolha de registro e de convenções da escrita.
40 pontos	Demonstra domínio precário da modalidade escrita formal da língua portuguesa, de forma sistemática, com diversificados e frequentes desvios gramaticais, de escolha de registro e de convenções da escrita.
0 ponto	Demonstra desconhecimento da modalidade escrita formal da língua portuguesa.

Tabela 2. Enem Competência 1

ANEXO C – AVALIAÇÃO DA REDAÇÃO DO ENEM – COMPETÊNCIA

2

COMPETÊNCIA 2

COMPREENDER A PROPOSTA DE REDAÇÃO E APLICAR CONCEITOS DAS VÁRIAS ÁREAS DE CONHECIMENTO PARA DESENVOLVER O TEMA, DENTRO DOS LIMITES ESTRUTURAIS DO TEXTO DISSERTATIVO-ARGUMENTATIVO EM PROSA

200 pontos	Desenvolve o tema por meio de argumentação consistente, a partir de um repertório sociocultural produtivo e apresenta excelente domínio do texto dissertativo-argumentativo.
160 pontos	Desenvolve o tema por meio de argumentação consistente e apresenta bom domínio do texto dissertativo-argumentativo, com proposição, argumentação e conclusão.
120 pontos	Desenvolve o tema por meio de argumentação previsível e apresenta domínio mediano do texto dissertativo-argumentativo, com proposição, argumentação e conclusão.
80 pontos	Desenvolve o tema recorrendo à cópia de trechos dos textos motivadores ou apresenta domínio insuficiente do texto dissertativo-argumentativo, não atendendo à estrutura com proposição, argumentação e conclusão.
40 pontos	Apresenta o assunto, tangenciando o tema, ou demonstra domínio precário do texto dissertativo-argumentativo, com traços constantes de outros tipos textuais.
0 ponto	Fuga ao tema/não atendimento à estrutura dissertativo-argumentativa. Nestes casos a redação recebe nota zero e é anulada.

Tabela 3. Enem Competência 2

ANEXO D – AVALIAÇÃO DA REDAÇÃO DO ENEM – COMPETÊNCIA

3

COMPETÊNCIA 3

SELECIONAR, RELACIONAR, ORGANIZAR E INTERPRETAR INFORMAÇÕES, FATOS, OPINIÕES E ARGUMENTOS EM DEFESA DE UM PONTO DE VISTA

200 pontos	Apresenta informações, fatos e opiniões relacionados ao tema proposto, de forma consistente e organizada, configurando autoria, em defesa de um ponto de vista.
160 pontos	Apresenta informações, fatos e opiniões relacionados ao tema, de forma organizada, com indícios de autoria, em defesa de um ponto de vista.
120 pontos	Apresenta informações, fatos e opiniões relacionados ao tema, limitados aos argumentos dos textos motivadores e pouco organizados, em defesa de um ponto de vista.
80 pontos	Apresenta informações, fatos e opiniões relacionados ao tema, mas desorganizados ou contraditórios e limitados aos argumentos dos textos motivadores, em defesa de um ponto de vista.
40 pontos	Apresenta informações, fatos e opiniões pouco relacionados ao tema ou incoerentes e sem defesa de um ponto de vista.
0 ponto	Apresenta informações, fatos e opiniões não relacionados ao tema e sem defesa de um ponto de vista.

Tabela 4. Enem Competência 3

ANEXO E – AVALIAÇÃO DA REDAÇÃO DO ENEM – COMPETÊNCIA

4

COMPETÊNCIA 4

DEMONSTRAR CONHECIMENTO DOS MECANISMOS LINGÜÍSTICOS NECESSÁRIOS PARA A CONSTRUÇÃO DA ARGUMENTAÇÃO

200 pontos	Articula bem as partes do texto e apresenta repertório diversificado de recursos coesivos.
160 pontos	Articula as partes do texto, com poucas inadequações, e apresenta repertório diversificado de recursos coesivos.
120 pontos	Articula as partes do texto, de forma mediana, com inadequações, e apresenta repertório pouco diversificado de recursos coesivos.
80 pontos	Articula as partes do texto, de forma insuficiente, com muitas inadequações e apresenta repertório limitado de recursos coesivos.
40 pontos	Articula as partes do texto de forma precária.
0 ponto	Não articula as informações.

Tabela 5. Enem Competência 4

ANEXO F – AVALIAÇÃO DA REDAÇÃO DO ENEM – COMPETÊNCIA

5

COMPETÊNCIA 5

ELABORAR PROPOSTA DE INTERVENÇÃO PARA O PROBLEMA ABORDADO, RESPEITANDO OS DIREITOS HUMANOS

200 pontos	Elabora muito bem proposta de intervenção, detalhada, relacionada ao tema e articulada à discussão desenvolvida no texto.
160 pontos	Elabora bem proposta de intervenção relacionada ao tema e articulada à discussão desenvolvida no texto.
120 pontos	Elabora, de forma mediana, proposta de intervenção relacionada ao tema e articulada à discussão desenvolvida no texto.
80 pontos	Elabora, de forma insuficiente, proposta de intervenção relacionada ao tema, ou não articulada com a discussão desenvolvida no texto.
40 pontos	Apresenta proposta de intervenção vaga, precária ou relacionada apenas ao assunto.
0 ponto	Não apresenta proposta de intervenção ou apresenta proposta não relacionada ao tema ou ao assunto.

Tabela 6. Enem Competência 5

ANEXO G

Apêndices

APÊNDICE A – DOCUMENTAÇÃO PRIMEIRA PARTE

In [1]:

```
TEXT = "../../../src/spekuloom/Textos Alunos.txt"
OUTPUT_DIR = "/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/"
def mangle_name(nomes):
    nome = (nomes[0][1:8].lower() + "_" + "_".join(a[:4//len(nomes)].lower()
for a in nomes[1:3])).ljust(12) + nomes[0][0]
    nome = nome
    nome = nome.translate(str.maketrans(" .Ã;Ã£ÃçÃ©ÃªÃ³Ã´ÃµÃ°",
"__aaaeioou"))
    return nome
with open(TEXT, "r") as textfile:
    text = textfile.read()
    texts = text.split("@")[1:]
    only_texts = ["\n".join(t.split("\n")[1:]) for t in texts]
    only_texts_heads = [mangle_name(t.split("\n")[0].split())
                        for t in texts]
    heads_and_texts = zip(only_texts_heads, only_texts)
    #[print(h, "\t", len(t)) for h, t in heads_and_texts]

    for head, out_text in heads_and_texts:
        filename = OUTPUT_DIR+f"{head}.txt"
        if len(out_text) <700:
            continue
        with open(filename, "w") as out_file:
            print(filename)
            out_file.write(out_text)
    ...

/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/g_s_k_0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/g_c_m_0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/j_s_p__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/j_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/l_s_k__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/j_b_t__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/g_d__0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/i_s_b_1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/j_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_n_p_1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/k_k_v_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/g_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_m_f__0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/c_m__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/r_c__0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/d_d_c__0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_c_a__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/j_d_e__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/y_a_d__0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/s_____1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/v_m__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/r_n_p__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/s_____1.txt
```

```

/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_c_d__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_c_b__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_a_m__1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/d_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/b_____1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/a_h_m__0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/c_____1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/a_l_f_____1.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/p_r_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_r_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/p_____0.txt
/home/carlo/Documentos/dev/spekuloom/src/SpekuloomS/basico/m_c_v__1.txt

```

In [2]:

```

import sys
SPEKULOOM_PATH = ".././src/"
sys.path.append(SPEKULOOM_PATH)
from spekuloom.core import Inscription, W, Z, Y
Inscription(outtable=SPEKULOOM_PATH+"spekuloom/samples.tab",
parameters=W).arrange_data_for_learning()
corpora0 34 [(10, 6), (5, 10), (5, 11), (12, 9), (8, 10), (8, 9), (6, 10),
(5, 8), (2, 6), (8, 9), (10, 7), (4, 13), (11, 7), (8, 11), (9, 9), (9, 8),
(10, 11), (7, 8), (11, 8), (6, 11), (10, 7), (7, 9), (6, 8), (10, 11), (10,
11), (9, 8), (10, 8), (11, 7), (12, 7), (5, 11), (6, 9), (5, 10), (8, 10),
(4, 11)]
patterns â-îâ-²â-îâ-² [2, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 2, 0, 1]
patterns â-□â-´â-²â-î [1, 1, 0, 0, 0, 2, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1,
0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]
patterns â-´â-²â-²â-² [0, 2, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 1, 0, 0, 0, 2, 0, 0, 0]
patterns â-□â-□â-´â-² [0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0,
0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
patterns â-□â-´â-²â-» [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-»â-□â-»â-² [0, 0, 2, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0]
patterns â-²â-□â-»â-□ [0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0]
patterns â-□â-»â-²â-¬ [0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-´â-²â-□â-¬ [0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2,
0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]
patterns â-»â-«â-□â-□ [0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-»â-□â-´â-² [0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 2,
0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1]
patterns â-«â-□â-»â-² [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-¬â-»â-□â-´ [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-□â-îâ-´â-² [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0,
0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0]
patterns â-²â-□â-´â-² [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-«â-□â-³â-² [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-²â-¬â-´â-² [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

```

```
['n', 'g', 'â-îâ-²â-îâ-²', 'â-□â-²â-²â-î', 'â-²â-²â-²â-²', 'â-□â-□â-²â-²',  
'â-□â-²â-²â-²», 'â-²â-²â-²â-²», 'â-²â-²â-²â-²', 'â-²â-²â-²â-²», 'â-²â-²â-²â-²-  
□â-²», 'â-²â-²â-²â-²», 'â-²â-²â-²â-²», 'â-²â-²â-²â-²», 'â-²â-²â-²â-²», 'â-²â-²â-²â-²»,  
'c']  
['s', 's', 'c', 'c', 'c', 'c', 'c', 'c', 'c', 'c', 'c', 'c', 'c', 'c', 'c', 'c', 'c',  
'c', 'c', 'c', 'c', 'd']  
['m', 'm', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ',  
' ', ' ', ' ', ' ', ' ', ' ', 'c']  
['d_____ ', '0', 2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['p_____ ', '0', 0, 1, 2, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['m_c_a___ ', '1', 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['v_m___ ', '1', 2, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 'basico']  
['c_m___ ', '1', 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0, 0, 0, 0, 0, 0, 'basico']  
['g_s_k___ ', '0', 0, 2, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0,  
'basico']  
['p_r_____ ', '0', 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['m_n_p___ ', '1', 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,  
'basico']  
['d_d_c_____ ', '0', 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,  
'basico']  
['g_c_m___ ', '0', 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0, 0,  
'basico']  
['c_____ ', '1', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['g_____ ', '0', 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['m_c_v_____ ', '1', 0, 0, 0, 2, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0,  
'basico']  
['j_d_e_____ ', '1', 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 1, 0, 0, 0, 0, 0, 0,  
'basico']  
['j_b_t_____ ', '1', 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0,  
'basico']  
['l_s_k_____ ', '1', 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['j_s_p_____ ', '1', 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0,  
'basico']  
['k_k_v_____ ', '0', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,  
'basico']  
['g_d_o_____ ', '0', 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 2, 0, 0,  
'basico']  
['m_____ ', '0', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['a_h_m_____ ', '0', 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0,  
'basico']  
['s_____ ', '1', 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,  
'basico']  
['j_p_____ ', '0', 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
'basico']  
['i_s_b_____', '1', 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 1, 1, 0, 0,  
'basico']  
['r_n_p_____ ', '1', 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1,  
'basico']  
['a_l_f_____ ', '1', 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,  
'basico']  
['s_____ ', '1', 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,  
'basico']
```

```

['m_re__', '0', 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
'basico']
['y_a_d____', '0', 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
'basico']
['m_m_f____', '0', 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
'basico']
['b____', '1', 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0,
'basico']
['m____', '0', 2, 0, 2, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 1, 0, 0, 2,
'basico']
['m_c_d____', '1', 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
'basico']
['r_c____', '0', 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
'basico']

```

In [3]:

```

Y.TXT_CUT = 2000
Inscription(outtable=SPEKULOOM_PATH+"spekuloom/learn20.tab",
parameters=Y).arrange_data_for_learning()
Y.TXT_CUT = 2400
Inscription(outtable=SPEKULOOM_PATH+"spekuloom/learn24.tab",
parameters=Y).arrange_data_for_learning()
Y.TXT_CUT = 4800
Inscription(outtable=SPEKULOOM_PATH+"spekuloom/learn48.tab",
parameters=Y).arrange_data_for_learning()
corpora0 33 [(15, 9), (16, 15), (8, 7), (11, 13), (17, 9), (20, 15), (18,
9), (0, 0), (5, 2), (0, 0), (20, 13), (16, 11), (13, 18), (57, 18), (38,
11), (9, 15), (7, 12), (36, 12), (26, 10), (9, 17), (16, 15), (6, 11), (16,
10), (14, 16), (21, 15), (14, 16), (7, 11), (3, 17), (11, 9), (14, 7), (17,
9), (14, 12), (21, 13)]
patterns â-â-â-â-â-â [1, 1, 0, 0, 0, 1, 2, 0, 0, 0, 1, 1, 2, 0, 0, 0, 0,
0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 1, 0]
patterns â-â-â-â-â-â [2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0,
0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0]
patterns â-â-â-â-â-â [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 0,
0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0]
patterns â-â-â-â-â-â [1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-â-â-â-â-â [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0]
patterns â-â-â-â-â-â [0, 1, 0, 1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 1, 2, 1,
1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 4, 0, 1, 1, 1, 0]
patterns â-â-â-â-â-â [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 2, 0, 2, 1, 0, 0, 1, 0, 0, 0]
patterns â-â-â-â-â-â [0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0,
0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-â-â-â-â-â [0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0]
patterns â-â-â-â-â-â [0, 0, 0, 1, 5, 1, 2, 0, 0, 0, 0, 1, 0]
patterns â-â-â-â-â-â [0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 2, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0]
patterns â-â-â-â-â-â [0, 0, 0, 1, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
patterns â-â-â-â-â-â [0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-â-â-â-â-â [0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0]
patterns â-â-â-â-â-â [0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

```



```
[ 'jose_da_roch', 'a', 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 'intermediario']
[ 'maraes_silva', ']', 0, 0, 0, 0, 0, 1, 0, 0, 0, 6, 0, 1, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 2, 0, 0, 'intermediario']
[ 'ois_principe', 's', 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 3, 0, 0, 0, 'intermediario']
[ 'es_guerreiro', 's', 0, 0, 0, 0, 0, 1, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 'intermediario']
[ 'ho_no_espelh', 'o', 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
0, 2, 0, 0, 1, 0, 0, 0, 0, 'intermediario']
[ 'ose_guimarae', 's', 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 'intermediario']
[ 'hermenegildo', '-', 2, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0,
1, 0, 0, 2, 0, 0, 0, 0, 'intermediario']
[ 'a-no-Interio', 'r', 0, 1, 0, 0, 0, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 1, 0, 0, 2, 0, 'intermediario']
[ 'stavo-iriart', 'e', 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0,
1, 0, 0, 2, 0, 0, 1, 0, 0, 'intermediario']
[ 'tarcisio-lag', 'e', 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 'transitorio']
[ 'n-machado.pd', 'f', 0, 0, 0, 0, 0, 1, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 'transitorio']
[ 'e-luiz-da-lu', 'z', 0, 0, 0, 0, 0, 0, 0, 0, 1, 5, 1, 1, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 1, 0, 2, 'transitorio']
[ '-leon-machad', 'o', 0, 0, 0, 0, 0, 1, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 'transitorio']
[ 'redo-pimente', 'l', 0, 0, 0, 0, 0, 0, 1, 0, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 1, 0, 0, 0, 0, 0, 'transitorio']
[ '-souza-vieir', 'a', 2, 0, 0, 0, 0, 2, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 'transitorio']
[ 'andre-compar', 't', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 0, 0, 1, 0, 2, 0, 1, 'transitorio']
[ '-almeida-hec', 'k', 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 2, 0, 1, 1, 0, 'transitorio']
[ '-almeida-hec', 'k', 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 2, 0, 1, 0, 0, 'transitorio']
[ '-teresa-lope', 's', 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
2, 0, 0, 1, 0, 0, 0, 0, 0, 'transitorio']
[ 'o-abel-sidne', 'y', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1,
0, 0, 1, 0, 0, 1, 0, 0, 0, 'transitorio']
corpora0 33 [(19, 14), (27, 20), (8, 7), (11, 13), (21, 14), (25, 20), (18,
9), (0, 0), (5, 2), (0, 0), (20, 13), (21, 18), (17, 22), (73, 23), (49,
14), (11, 20), (9, 19), (46, 17), (38, 15), (17, 21), (21, 20), (9, 16),
(21, 14), (19, 21), (24, 20), (19, 21), (12, 15), (5, 24), (17, 13), (18,
13), (24, 13), (16, 17), (29, 18)]
patterns â-â-â-â-â- [1, 2, 0, 0, 0, 0, 1, 2, 0, 0, 0, 1, 1, 2, 0, 0, 0, 1,
1, 0, 4, 0, 0, 2, 0, 0, 0, 1, 2, 0, 1, 0, 1, 0]
patterns â-â-â-â-â- [2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 2, 0, 0, 1,
0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
patterns â-â-â-â-â- [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 0,
0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0]
patterns â-â-â-â-â- [1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 0, 1, 1, 0, 2, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0]
patterns â-â-â-â-â- [1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-â-â-â-â- [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0]
patterns â-â-â-â-â- [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 1, 0,
2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
patterns â-â-â-â-â- [0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0]
```



```

['-lopes-vieir', 'a', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'basico']
['ona-Mulazzan', 'i', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'basico']
['ylvia-Manzan', 'o', 1, 2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,
2, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'basico']
['-queria-a-lu', 'a', 1, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0,
2, 2, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 2, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'intermediario']
['san-pedrimbi', '-', 2, 0, 2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1,
1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0,
0, 0, 0, 'intermediario']
['jose_da_roch', 'a', 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'intermediario']
['maraes_silva', ']', 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
7, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 2, 0, 0,
0, 0, 0, 'intermediario']
['ois_principe', 's', 0, 0, 0, 0, 0, 0, 1, 0, 0, 3, 0, 0, 0, 0, 0, 0, 2, 0, 1,
0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 2, 2, 0, 2, 0,
0, 0, 0, 'intermediario']
['es_guerreiro', 's', 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0,
2, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'intermediario']
['ho_no_espelh', 'o', 1, 0, 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'intermediario']
['ose_guiimarae', 's', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 0, 0, 'intermediario']
['hermenegildo', '-', 4, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 1, 0, 0,
1, 2, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'intermediario']
['a-no-Interio', 'r', 0, 1, 0, 1, 0, 1, 0, 0, 0, 2, 1, 0, 0, 0, 0, 1, 2, 0,
2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 0, 0, 0, 1, 0, 1, 2,
0, 0, 0, 'intermediario']
['stavo-iriart', 'e', 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0,
0, 0, 0, 2, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 2, 0, 0, 0, 1, 0, 0,
2, 0, 0, 'intermediario']
['tarcisio-lag', 'e', 2, 0, 1, 2, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 0, 0, 'transitorio']
['n-machado.pd', 'f', 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
1, 0, 0, 'transitorio']
['e-luiz-da-lu', 'z', 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1,
5, 1, 1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
0, 2, 0, 'transitorio']
['-leon-machad', 'o', 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0,
1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
1, 0, 0, 'transitorio']
['redo-pimente', 'l', 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 2,
2, 0, 0, 0, 0, 0, 0, 0, 1, 2, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 'transitorio']
['-souza-vieir', 'a', 2, 0, 0, 0, 0, 0, 0, 2, 0, 5, 0, 0, 0, 0, 0, 2, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0,
1, 0, 2, 'transitorio']

```


patterns $\hat{a}-\hat{a}'\hat{a}-\hat{a}^2\hat{a}-\hat{a}$ [2, 2, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1,
 0, 0, 1, 1, 0, 2, 1, 2, 1, 1, 3, 0, 0, 1, 1, 2]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}\rightarrow\hat{a}^2$ [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0,
 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 2, 0, 1, 1, 0]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}'\hat{a}^2$ [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 1, 0,
 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0]
 patterns $\hat{a}^2\hat{a}-\hat{a}'\hat{a}^2\hat{a}-\hat{a}$ [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
 patterns $\hat{a}-\hat{a}'\hat{a}^2\hat{a}-\hat{a}\hat{a}$ [1, 0, 0, 0, 0, 2, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1,
 1, 2, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 2, 0]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}'\hat{a}-\hat{a}'$ [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
 0, 0, 2, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0]
 patterns $\hat{a}-\hat{a}'\hat{a}^2\hat{a}\rightarrow\hat{a}^2$ [1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2,
 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
 patterns $\hat{a}-\hat{a}\hat{a}'\hat{a}\rightarrow\hat{a}^2$ [1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 0, 0, 1, 0,
 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1]
 patterns $\hat{a}-\hat{a}'\hat{a}^2\hat{a}-\hat{a}\hat{a}$ [1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0, 3, 0,
 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1]
 patterns $\hat{a}^2\hat{a}-\hat{a}\hat{a}'\hat{a}^2$ [1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 1, 3, 1, 0, 2, 1, 2, 0, 0, 0, 0, 0, 0, 0]
 patterns $\hat{a}-\hat{a}\hat{a}^2\hat{a}\rightarrow\hat{a}\hat{a}$ [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 0]
 patterns $\hat{a}^2\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}^2$ [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0,
 2, 1, 0, 2, 0, 0, 0, 0, 0, 2, 2, 0, 0, 1, 2, 0]
 patterns $\hat{a}-\hat{a}'\hat{a}^2\hat{a}-\hat{a}\hat{a}$ [0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0]
 patterns $\hat{a}^2\hat{a}-\hat{a}\hat{a}^2\hat{a}-\hat{a}$ [0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 2, 2,
 3, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}$ [0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0]
 patterns $\hat{a}^2\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}$ [0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 2,
 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}\rightarrow\hat{a}^2$ [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 2, 0, 0, 3, 0, 0, 0, 0, 0, 0, 0, 2, 2, 0, 0, 1]
 patterns $\hat{a}-\hat{a}\hat{a}\rightarrow\hat{a}^2\hat{a}-\hat{a}$ [0, 3, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 0, 1, 1, 0, 1, 0]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}'\hat{a}^2$ [0, 3, 0, 0, 0, 0, 1, 0, 0, 0, 1, 2, 2, 1, 0, 0, 2,
 1, 0, 1, 0, 1, 1, 0, 3, 0, 0, 2, 0, 2, 2, 0, 0]
 patterns $\hat{a}-\hat{a}\hat{a}\rightarrow\hat{a}-\hat{a}\hat{a}$ [0, 2, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}^2\hat{a}-\hat{a}$ [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
 patterns $\hat{a}\rightarrow\hat{a}-\hat{a}'\hat{a}^2$ [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
 1, 0, 0, 0, 0, 3, 0, 0, 0, 1, 1, 2, 0, 1, 2]
 patterns $\hat{a}\rightarrow\hat{a}^2\hat{a}-\hat{a}\hat{a}$ [0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
 patterns $\hat{a}-\hat{a}\hat{a}^2\hat{a}-\hat{a}$ [0, 2, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0]
 patterns $\hat{a}\rightarrow\hat{a}-\hat{a}\hat{a}-\hat{a}$ [0, 2, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1,
 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}\hat{a}-\hat{a}^2$ [0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1,
 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
 patterns $\hat{a}-\hat{a}'\hat{a}^2\hat{a}-\hat{a}\hat{a}$ [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
 0, 0, 0, 0, 2, 0, 0, 2, 0, 1, 0, 0, 0, 0, 0, 0]
 patterns $\hat{a}-\hat{a}\hat{a}^2\hat{a}-\hat{a}^2$ [0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 2, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]
 patterns $\hat{a}-\hat{a}\hat{a}-\hat{a}'\hat{a}^2$ [0, 2, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 2,
 1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0]

0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 7, 0, 0, 0, 1, 0, 0, 1, 1, 1,
 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 2, 0, 0, 0,
 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 'intermediario']
 ['ois_principe', 's', 0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 1, 0, 0, 0, 3,
 2, 0, 0, 0, 1, 0, 0, 1, 0, 1, 3, 0, 0, 0, 0, 2, 0, 1, 2, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 1, 0, 0, 2, 0, 0, 0, 2, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 3, 0, 0, 2, 1, 2, 1, 1, 1, 0, 1, 0, 0, 0,
 0, 1, 2, 0, 0, 1, 0, 2, 0, 0, 0, 0, 0, 0, 1, 0, 4, 0, 0, 0, 2, 0, 0, 0,
 0, 0, 0, 1, 0, 0, 2, 0, 1, 0, 0, 0, 0, 0, 3, 2, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 'intermediario']
 ['es_guerreiro', 's', 1, 0, 2, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 2,
 1, 0, 1, 0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0, 2, 0, 2, 1, 0, 0, 2, 0, 0, 0,
 0, 0, 1, 1, 0, 0, 2, 0, 0, 1, 0, 0, 0, 0, 1, 2, 1, 0, 0, 0, 0, 1, 0, 0, 1,
 0, 0, 1, 0, 0, 0, 0, 2, 0, 0, 0, 0, 2, 0, 1, 0, 0, 0, 0, 0, 2, 0, 0, 2,
 1, 2, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0,
 1, 0, 0, 0, 3, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 'intermediario']
 ['ho_no_espelh', 'o', 1, 2, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 2, 0, 0, 0, 1,
 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 2, 0, 3, 0, 0, 0, 2, 0, 1, 0, 0, 1,
 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 2, 0, 2, 3, 0, 0, 0, 0, 0, 0, 0, 1,
 0, 0, 1, 0, 0, 1, 0, 0, 0, 3, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 2, 1, 0, 2,
 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 1, 1, 1, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 'intermediario']
 ['ose_guimarae', 's', 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 2, 1, 0, 0, 0, 0, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 1, 0, 0, 0,
 0, 0, 0, 2, 0, 1, 1, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 2, 1, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 3,
 1, 1, 2, 0, 0, 0, 0, 0, 0, 'intermediario']
 ['hermenegildo', '-', 4, 1, 1, 0, 2, 0, 2, 1, 2, 1, 0, 0, 0, 1, 1, 0, 0, 1,
 1, 0, 1, 1, 1, 0, 0, 2, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 0, 3, 2, 0, 2, 3, 0, 1, 0, 0, 0,
 1, 2, 2, 0, 2, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 1, 1, 2, 0, 2, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0, 0, 1, 2, 0, 0, 0,
 0, 0, 0, 2, 1, 1, 0, 0, 0, 0, 'intermediario']
 ['a-no-Interio', 'r', 2, 0, 3, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 3,
 2, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 3, 0, 2, 0, 1, 0, 0, 0, 3, 0, 0, 0, 0, 0,
 0, 0, 0, 1, 0, 0, 2, 1, 0, 0, 0, 0, 0, 1, 0, 7, 0, 0, 0, 0, 2, 0, 0, 2, 0,
 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 3, 0, 0, 0, 2, 0, 0, 0,
 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 2, 2, 3, 0, 0, 2, 0, 0, 0, 3, 0, 0, 0,
 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 2, 0, 0, 1, 2, 0, 0, 1, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 'intermediario']
 ['stavo-iriart', 'e', 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1,
 1, 2, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
 0, 0, 0, 0, 2, 0, 0, 1, 2, 0, 0, 0, 0, 0, 1, 3, 1, 0, 0, 0, 1, 1, 0, 0, 1,
 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 3, 0,
 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 2, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
 0, 1, 0, 0, 0, 3, 0, 0, 0, 0, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 'intermediario']
 ['tarcisio-lag', 'e', 3, 0, 0, 1, 0, 0, 2, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0, 1,
 3, 1, 2, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 3,
 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 2, 0, 0, 2, 4, 0, 0, 0, 0, 0, 1, 0, 0, 0,
 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0,
 0, 1, 2, 0, 0, 0, 0, 0, 0, 2, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 2, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 'transitorio']

```

['n-machado.pd', 'f', 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1,
1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0,
0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 3, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 2, 0, 1, 0, 2, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,
2, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 2, 0, 0, 'transitorio']
['e-luiz-da-lu', 'z', 1, 0, 2, 0, 0, 1, 0, 0, 3, 0, 1, 0, 0, 0, 0, 1, 0, 0,
0, 0, 2, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 2, 3, 0, 0, 0,
0, 0, 1, 0, 2, 0, 0, 0, 0, 0, 1, 0, 0, 2, 0, 5, 1, 0, 0, 0, 1, 1, 1, 0, 0,
0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 2, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0,
0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
2, 0, 0, 0, 2, 0, 0, 0, 0, 'transitorio']
['-leon-machad', 'o', 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1,
1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0,
0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 3, 0, 0, 0, 1, 0, 1, 0, 0, 0,
0, 0, 0, 0, 0, 2, 0, 1, 0, 2, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 2, 0, 0, 0, 0, 0, 1, 1, 0, 0,
2, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 2, 0, 0, 'transitorio']
['redo-pimente', 'l', 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
2, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0,
1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 2, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 1,
1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0,
1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 'transitorio']
['-souza-vieir', 'a', 4, 0, 3, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 9,
3, 0, 3, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 2, 3, 1, 0, 0, 0, 0, 0, 2, 0, 0, 1,
0, 1, 0, 0, 0, 1, 2, 3, 0, 1, 1, 0, 1, 0, 1, 2, 0, 2, 0, 1, 1, 2, 2, 0, 1,
0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 2, 2, 0, 2, 2, 0,
1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0,
0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 2, 0, 1, 0, 0, 2, 0, 1, 0, 0, 0, 0, 1,
0, 1, 1, 0, 0, 0, 3, 0, 1, 0, 'transitorio']
['andre-compar', 't', 2, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 2, 1, 0, 0, 1, 0,
0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0, 0, 1,
0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
0, 2, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 1, 0, 0, 2, 3, 'transitorio']
['-almeida-hec', 'k', 3, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 0, 1, 0, 0, 0, 5,
1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 1, 2, 0, 0, 2,
0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 2, 0, 0, 0, 0, 0, 0, 1, 0, 1,
0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 2, 1, 0, 0, 2, 0, 0, 0,
0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 2, 0, 1, 0, 1, 0, 0, 2, 0, 0, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1,
0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 'transitorio']
['-almeida-hec', 'k', 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2,
3, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 2, 0, 0, 0, 0, 2, 0, 0, 0,
0, 0, 0, 0, 0, 0, 1, 1, 0, 3, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0,
2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 1, 0, 0, 0, 2, 0, 0, 0,
0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 2, 0,
0, 0, 0, 0, 0, 0, 1, 1, 0, 'transitorio']
['-teresa-lope', 's', 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 2,
5, 0, 1, 1, 0, 0, 2, 0, 0, 0, 1, 0, 2, 2, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1,
0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0,
0, 1, 0, 0, 0, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 5, 0,

```



```
0, 0, 0, 0, 0, '2'], ['m_c_d____', '1', '0', 0, 0, 0, 0, 0, 0, 0, 0, '0', 0, '0',
0, 0, '0', 0, 0, 0, 0, 0, 0, 0, 0, 0, '0', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, '1', 0, 0, 0, 0, 0, 0, 0, '0'], ['r_c____', '0', '0', 0, 0, 0,
0, 0, 0, '0', 0, '1', 0, 0, '0', 0, 0, 0, 0, 0, 0, 0, 0, 0, '0', 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, '0', 0, 0, 0, 0, 0, 0, '0']]
```

In [17]:

```
import pandas as pd
df = pd.read_csv(SPEKULOOM_PATH+"spekuloom/predict24.tab", delimiter='\t',
header=[0], skiprows=[1,2])
df
df1 = df[df.columns[0:4]] #
dfs = df1.sort_values(by=['bsc', 'itm', 'trs'])
ax = dfs.plot.bar(stacked=True, figsize=(16,8))
_ = ax.set_xticklabels(dfs[dfs.columns[0]], rotation=90)
```

In [54]:

```
import pandas as pd
df = pd.read_csv("gram.tab", delimiter='\t')
df
```

Out[54]:

	patt	ba	it	tr
0	â-â-²â—jâ-²	11	72	43
1	â-²â—jâ-²â-²	4	13	10
2	â—jâ-²â—jâ-²	16	31	29
3	â-²â—jâ-²â-²	3	19	11
4	â-²â—jâ-²â—jâ-²	6	15	11
5	â-²â-²â-²â—□	9	1	6
6	â-²â-²â-²â—□	3	16	10
7	â—□â-²â-²â—j	13	30	37
8	â—□â—jâ-²â-²	8	16	19
9	â—□â—□â-²â-²	5	10	11
10	â-²â-²â—□â—j	9	16	14
11	â—□â—jâ-²â—j	4	21	11
12	â—jâ-²â-²â—j	10	13	17
13	â-²â—□â-²â-²	3	12	3

	patt	ba	it	tr
14	$\hat{a} \rightarrow \hat{a} \square \hat{a} \rightarrow \hat{a}'$	2	3	5
15	$\hat{a} \rightarrow \hat{a} \square \hat{a}' \hat{a}^{-2}$	3	11	12
16	$\hat{a}' \hat{a}^{-2} \hat{a} \square \hat{a}'$	2	15	9
17	$\hat{a}' \hat{a}^{-2} \hat{a} \rightarrow \hat{a} \square$	5	11	14
18	$\hat{a}^{-2} \hat{a} \rightarrow \hat{a}' \hat{a}^{-2}$	2	6	8
19	$\hat{a} \square \hat{a} \rightarrow \hat{a} \rightarrow \hat{a}^{-2}$	3	19	4
20	$\hat{a}^{-2} \hat{a} \rightarrow \hat{a}^{-2} \hat{a} \rightarrow \hat{a} \rightarrow \hat{a}^{-2}$	6	23	26
21	$\hat{a}^{-2} \hat{a} \rightarrow \hat{a}^{-2} \hat{a} \rightarrow \hat{a} \rightarrow \hat{a}^{-2}$	1	14	8
22	$\hat{a}^{-2} \hat{a} \square \hat{a} \rightarrow \hat{a}^{-2}$	1	12	9
23	$\hat{a} \rightarrow \hat{a}' \hat{a}^{-2} \hat{a} \square$	2	5	7
24	$\hat{a}^{-2} \hat{a}^{-2} \hat{a} \rightarrow \hat{a}^{-2}$	0	17	8
25	$\hat{a}^{-2} \hat{a}' \hat{a}^{-2} \hat{a} \square$	0	10	3
26	$\hat{a}^{-2} \hat{a} \square \hat{a}' \hat{a}^{-2}$	0	22	18
27	$\hat{a} \rightarrow \hat{a}^{-2} \hat{a}^{-2} \hat{a}^{-2}$	0	7	1
28	$\hat{a} \rightarrow \hat{a}' \hat{a}^{-2} \hat{a}^{-2}$	0	7	1
29	$\hat{a}^{-2} \hat{a} \rightarrow \hat{a} \square \hat{a} \rightarrow$	0	3	10
30	$\hat{a}' \hat{a}^{-2} \hat{a}^{-2} \hat{a} \rightarrow \hat{a} \rightarrow \hat{a}^{-2}$	0	9	5
31	$\hat{a}' \hat{a}^{-2} \hat{a} \rightarrow \hat{a} \square$	0	7	0

Tabela 7. Primeira Tabela da Documentação de Avaliação Textual de Padrões
In [72]:

```
from matplotlib import pyplot as plt
# fig, px = plt.subplots(figsize=(6,5),dpi=200)

ax = df.plot.scatter(x="ba", y="it", s=df.tr*8, figsize=(16,8))
for i, txt in enumerate(df.patt):
    ax.annotate(txt, (df.ba.iat[i],df.it.iat[i]))
```

In [49]:

```
df.plot.scatter(x="ba", y="tr")
```

Out[49]:

<matplotlib.axes._subplots.AxesSubplot at 0x7fa3c17ea9b0>

In [50]:

```
df.plot.scatter(x="in", y="tr")
```

Out[50]:

<matplotlib.axes._subplots.AxesSubplot at 0x7fa3bdf5bbe0>

APÊNDICE B – DOCUMENTAÇÃO SEGUNDA PARTE

In [2]:

```
SPEKULOOM_PATH = "../..src/"
import pandas as pd
df = pd.read_csv(SPEKULOOM_PATH+"spekuloom/predict24.tab", delimiter='\t',
header=[0], skiprows=[1,2])
df1 = df[df.columns[0:4]] #
dfs = df1.sort_values(by=['bsc', 'itm', 'trs'])
ax = dfs.plot.bar(stacked=True, figsize=(16,8))
_ = ax.set_xticklabels(dfs[dfs.columns[0]], rotation=90)
```

In [5]:

```
df["cpx"] = [ row.bsc * 4 + row.itm * 7 + row.trs * 10 for index, row in
df.iterrows() ]
df.sort_values(by=['cpx'])
```

Out[5]:

	bsc	itm	trs	g	nbs	cpx
6	0.899158	0.048169	0.052673	0	0.0	4.460545
22	0.797707	0.085469	0.116825	0	0.0	4.957355
8	0.797707	0.085469	0.116825	0	0.0	4.957355
31	0.753682	0.080752	0.165566	1	0.0	5.235652
29	0.619452	0.199110	0.181439	0	0.0	5.685960
14	0.536903	0.345152	0.117945	0	0.0	5.743125
20	0.536903	0.345152	0.117945	0	0.0	5.743125
32	0.511501	0.383626	0.104874	0	0.0	5.780119
16	0.626090	0.080497	0.293413	1	0.0	6.001968
2	0.626090	0.080497	0.293413	1	0.0	6.001968
23	0.626090	0.080497	0.293413	0	0.0	6.001968

	bsc	itm	trs	g	nbs	cpx
24	0.626090	0.080497	0.293413	0	0.0	6.001968
9	0.545990	0.070199	0.383811	1	0.0	6.513465
1	0.407902	0.209778	0.382320	1	0.0	6.923256
27	0.411734	0.105875	0.482391	0	2.0	7.211970
28	0.411734	0.105875	0.482391	1	2.0	7.211970
21	0.411734	0.105875	0.482391	1	2.0	7.211970
26	0.411734	0.105875	0.482391	1	2.0	7.211970
19	0.411734	0.105875	0.482391	0	2.0	7.211970
0	0.411734	0.105875	0.482391	1	2.0	7.211970
15	0.411734	0.105875	0.482391	1	2.0	7.211970
10	0.411734	0.105875	0.48239	1	2.0	7.211970
7	0.411734	0.105875	0.482391	1	2.0	7.211970
5	0.411734	0.105875	0.482391	1	2.0	7.211970
18	0.411734	0.105875	0.482391	0	2.0	7.211970
11	0.225213	0.405383	0.369405	1	1.0	7.432577
3	0.232544	0.358782	0.408675	1	2.0	7.528393
33	0.232544	0.358782	0.408675	1	2.0	7.528393
25	0.190101	0.342181	0.467719	0	2.0	7.832854
4	0.193088	0.297908	0.509004	0	2.0	7.947745
17	0.196831	0.157465	0.645704	1	2.0	8.346619
13	0.196831	0.157465	0.645704	0	2.0	8.346619
30	0.196831	0.157465	0.645704	0	2.0	8.346619
12	0.148793	0.119034	0.732173	1	2.0	8.750139

Tabela 8. Segunda Tabela da Documentação de Avaliação Textual de Padrões
In [34]

```
import seaborn as sns
```

```
#df.cpx.groupby("g")
#sns.distplot(df.cpx)
```

```
sns.violinplot(x=df["g"], y=df.cpx)
```

Out[34]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f73c3b21c88>
```

In [35]:

```
sns.distplot(df.cpx)
```

Out[35]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f73c3948978>
```

In [53]:

```
sns.violinplot(hue="g", x= df.bsc, split=True, data=df)
```

Out[53]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f73c367f278>
```

In [55]:

```
sns.jointplot(x=df.bsc, y=df.trs, data=df)
```

Out[55]:

```
<seaborn.axisgrid.JointGrid at 0x7f73c35cd908>
```

In [59]:

```
sns.swarmplot(y=df.cpx, data=df,
              hue='g')
```

Out[59]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f73c3856a20>
```

In [61]:

```
melted_df = pd.melt(df,
                    id_vars=["n", "bsc", "itm", "trs"], # Variables to keep
                    var_name="Stat") # Name of melted variable
melted_df.head()
```

Out[61]:

	bsc	itm	trs	Stat	value
0	0.411734	0.105875	0.482391	g	1.0
1	0.407902	0.209778	0.382320	g	1.0

	bsc	itm	trs	Stat	value
2	0.626090	0.080497	0.293413	g	1.0
3	0.232544	0.358782	0.408675	g	1.0
4	0.193088	0.297908	0.509004	g	0.0

Tabela 9. Terceira Tabela da Documentação de Avaliação Textual de Padrões
In [66]:

```
ax = sns.violinplot(y="cpx", x="nbs", hue="g",  
...                 data=df, palette="muted", split=True)
```