**Relatório de anotação do corpus PetroNer — um corpus com anotações sintáticas e de entidades**

**Elvis de Souza, Maria Clara Castro, Tatiana Cavalcanti**

**Supervisão: Cláudia Freitas**

**Outubro de 2022**

**v.2 17 de maio 2023**

# Apresentação geral do corpus

O presente relatório se refere ao corpus PetroNer em sua versão do dia 20 de junho de 2023. O corpus foi construído a partir de um conjunto de 11 Boletins Técnicos da Petrobras, que foram pré-processados pelo Laboratório de Inteligência Computacional Aplicada (ICA/PUC-Rio), e apresenta anotação de entidades mencionadas do domínio do petróleo padrão ouro. Na tabela a seguir apresentamos as características do corpus.

| **documentos** | 11 boletins |
| --- | --- |
| **frases** | 24035 |
| **tokens** | 615418 |

O pré-processamento do PetroNer envolveu (1) a transformação dos arquivos, originalmente em PDF, para TXT, removendo conteúdos desnecessários para o processamento linguístico do PetroNer, como tabelas e figuras (ver mais em CAVALCANTI et al., 2021), e (2) a segmentação do arquivo TXT em frases e palavras (mais especificamente, tokens linguísticos), de acordo com as especificações do formato de anotação multilíngue Universal Dependencies. Realizamos ainda um pós-processamento para remover do corpus frases em inglês, procurando pelo token “the” e removendo as frases em que aparece. Como resultado, foram removidas 3.629 frases do corpus.

# Anotação sintática

## **Objetivo**

Para auxiliar no processo de anotação semântica dos boletins sem que fosse necessário ler e analisar todas as mais de 24 mil frases do corpus, utilizamos informações linguísticas de outros tipos, como a anotação de lemas (as formas de dicionário das palavras), classes gramaticais e, principalmente, a anotação sintática.

A lematização permite formular regras de anotação que dispensam as flexões das palavras – ao buscar pelo lema “bacia”, por exemplo, encontramos também as formas “Bacia”, “bacias”, “BACIA”, “Bacias”, etc. Em casos como “formação”, cujo plural é realizado trocando mais de uma letra, a anotação de lema é especialmente útil.

A anotação morfossintática é um outro ótimo atalho para encontrar entidades mencionadas, por exemplo quando há casos de coordenação e de aposição. Nas frases a seguir o termo em negrito está coordenado (1) ou está em estrutura de aposto (2) com o termo em itálico, indicando que são entidades do mesmo tipo ou de tipos relacionados, permitindo a formulação de regras inteiramente linguísticas para realizar a anotação semântica. Por serem apenas facilitadores, os resultados da aplicação dessas regras precisam ser e de fato foram revistos.

**coordenação:** Entre 1981 e 1990 fez parte de a equipe de avaliação de perfis e teste em as bacias de Campos e de o **Paraná** , dedicando se a a área de hidrodinâmica e hidroquímica . .

**aposto:** A maior acumulação de arenito asfáltico de a borda leste de a Bacia de o Paraná ( **Fazenda Betumita** ) ocorre em a área de estudo com um volume estimado de 5,7 milhões de bar-ris de óleo ( Thomaz Filho , 1982 ) . .

## **Metodologia**

A anotação morfossintática dos boletins foi realizada utilizando um modelo de anotação gerado pelo Stanza (Qi, 2020). O modelo foi treinado utilizando o dataset PetroGold v2 (de Souza, 2022), composto por 19 documentos acadêmicos do domínio do óleo & gás com anotação morfossintática padrão ouro.

Para gerar um bom modelo de anotação automática de dependências sintáticas, utilizamos o dataset inteiro para o treinamento, sem a separação entre partições de treino, desenvolvimento e teste. No lugar do particionamento, para avaliar a qualidade da anotação sintática dos boletins, realizamos a revisão de 50 frases do PetroNer anotadas pelo modelo, selecionadas de maneira semi-aleatória (selecionadas à mão por um dos anotadores) para comparar a análise automática com a análise correta.

Durante o processo de anotação semântica dos boletins, desenvolvemos algumas regras de revisão morfossintática para corrigir algumas frases que, porque estavam incorretamente anotadas no nível sintático, tiveram a anotação semântica afetada negativamente. Assim, uma vez que corrigimos essas frases, as regras de anotação semântica dependentes de informação morfossintática que já tínhamos desenvolvido foram corretamente aplicadas.

## **Resultados**

Para as 50 frases retiradas dos boletins e revistas pelos anotadores, o modelo treinado no PetroGold v2 obteve os seguintes resultados:

Metric | Precision | Recall | F1 Score | AligndAcc

-----------+-----------+-----------+-----------+-----------

Tokens | 100.00 | 100.00 | 100.00 |

Sentences | 100.00 | 100.00 | 100.00 |

Words | 100.00 | 100.00 | 100.00 |

**UPOS | 98.37 | 98.37 | 98.37 | 98.37**

XPOS | 100.00 | 100.00 | 100.00 | 100.00

UFeats | 99.22 | 99.22 | 99.22 | 99.22

AllTags | 98.19 | 98.19 | 98.19 | 98.19

**Lemmas | 98.67 | 98.67 | 98.67 | 98.67**

**UAS | 94.51 | 94.51 | 94.51 | 94.51**

**LAS | 92.58 | 92.58 | 92.58 | 92.58**

**CLAS | 88.28 | 87.23 | 87.75 | 87.23**

MLAS | 86.96 | 85.93 | 86.44 | 85.93

BLEX | 86.96 | 85.93 | 86.44 | 85.93

São resultados melhores que os da avaliação intrínseca do PetroGold v2 (quando se separaram as partições de treinamento, teste e desenvolvimento do corpus): em uma das métricas mais difíceis, o CLAS, o modelo treinado no corpus inteiro desempenhou 3,68 pp. melhor nos boletins do que na avaliação intrínseca do PetroGold.

Já as correções morfossintáticas dos boletins, decorrentes do processo de anotação semântica do corpus, afetaram 6336 tokens – seja na informação de lema, POS, relação de dependência ou governante da relação de dependência –, o que corresponde a 0,01% do número de tokens do corpus.

# Anotação semântica

## **O tagset e o esquema de anotação**

O tagset utilizado para a anotação semântica do corpus PetroNer contém 17 classes. Abaixo estão listadas todas as classes, que foram definidas através de uma ontologia criada pelo grupo de especialistas da Petrobras. Tanto o tagset quanto o esquema de anotação estão detalhados em Castro et al., 2022.

**BACIA**

**CAMPO**

**ESTRUTURA\_FÍSICA:** Estrutura física do objeto geológico

**EVENTO\_PETRO:** Eventos do sistema petrolífero

**FLUIDO**

**FLUIDODATERRA\_i:** Fluido da terra (de interesse da indústria do petróleo)

**FLUIDODATERRA\_o:** Fluido da terra (outros)

**NÃOCONSOLID:** Material não consolidado

**POÇO**

**POÇO\_Q:** Reclassificação de poços (qualidade)

**POÇO\_R:** Reclassificação de poços (role - papel)

**POÇO\_T:** Reclassificação de poços (tipo)

**ROCHA**

**TEXTURA:** Textura das rochas

**TIPO\_POROSIDADE:** Porosidade das rochas

**UNIDADE\_CRONO:** Unidade cronoestratigráfica

**UNIDADE\_LITO:** Unidade litoestratigráfica

## **O processo de anotação automática**

A anotação automática das entidades é realizada em etapas.

**(1)** Primeiro, anotamos os léxicos coletados pelo CENPES, onde procuramos pelos termos do léxico nos textos dos boletins.

O script `lexicons.py` realiza a anotação dos arquivos conllu realizando antes alguns pré-processamentos das entidades no léxico. O procedimento consiste em acionar o Stanza para tokenizar e lematizar os elementos do léxico, além de remover acentos gráficos e tornar as letras todas minúsculas. Assim, se uma entrada do léxico é “CAMPOS DO JORDÃO” (entrada fictícia), pouco importa se no texto do arquivo *conllu* a ocorrência é “campo de o Jordão” ou “campo do jordao”, ambas serão anotadas como do tipo especificado, sendo a primeira palavra do tipo “B” (begin) e as demais do tipo “I” (in). Caberá às regras de correção, mais à frente, corrigir estes casos específicos em que pode ter ocorrido algum erro.

Manipulamos algumas das lematizações automáticas realizadas pelo Stanza que geravam muitos erros na anotação semântica, conforme se lê no dicionário abaixo (lê-se: quando a palavra é a da esquerda, seu lema deve ser o da direita – assim, o lema de “santos” deve ser “santos”, e não “santo”, como seria de se esperar, para evitar erros de atribuição de entidade).

lemmas = {

'ums': 'ums',

'santos': 'santos',

'estai': 'estai',

'veio': 'veio'

}

**(2)** Em seguida, acionamos as regras de revisão e de anotação semântica, cujo código está disponível em `rules.py`. As regras estão divididas em alguns blocos por tipo de anotação, e para cada bloco é feita uma nova iteração em todo o corpus, tornando mais difícil assim que alguma regra deixe de ser aplicada por conta da ordem em que se apresentam no código.

Os blocos são os seguintes:

**(2.1)** Correções sintáticas – são as primeiras, pois algumas das regras de anotação semântica dependem dos outros níveis linguísticos

**Exemplo:**

if sentid == "boletins-000008-1528" and token.id == "85":

reattach(token, "78")

if sentid == "boletins-000011-433" and token.id == "44":

token.deprel = "nmod:appos"

**(2.2)** Léxico via regra – o bloco destina-se às palavras que, embora não tenham sido reunidas nos léxicos do CENPES, devem ser anotadas segundo discussões internas (devidamente documentadas).

**Exemplo:**

if regex(“Cedro|Tucano|Jatobá|Recôncavo-Tucano-Jatobá”, token.lemma):

token.deps = "B=BACIA"

if regex("petróleo", token.lemma):

token.deps = "B=FLUIDODATERRA"

**(2.3)** Regras de expansão – essas regras têm como objetivo procurar outras entidades que, pela estrutura linguística, provavelmente são de fato entidades, mas não estavam nos léxicos.

**Exemplo:**

# trigger – regras de expansão para entidades que têm palavras “gatilho”, como “bacia de X”, “campo X”, “formação X”, etc.

[trigger] (de) (o) PROPN

B= I= I= I=

ou

O O O B=

# conj – para termos coordenados a entidades, que portanto também devem ser entidades

Tucano Sul e Norte

B= I= O B=

# flat:name – para termos anotados como “flat:name” sintaticamente, isto é, nomes próprios compostos, quando um dos termos do nome próprio já foi identificado como entidade

antes: Bacia\_B de\_I o\_I Rio\_I Almada\_O

depois: Bacia\_B de\_I o\_I Rio\_I Almada\_I

**(2.4)** Regras de revisão – têm como objetivo limpar erros que possam ter ocorrido devido à anotação inicial dos léxicos ou mesmo de outras regras anteriores

**Exemplo:**

if regex("Península|Atlântico|Nordeste|Brasil|Oriente", token.lemma) and regex("B=BACIA", token.deps):

token.deps = "O"

if token.word == "Rio" and token.next\_token.word == "de" and token.next\_token.next\_token.word == "Janeiro":

token.deps = "O"

**(2.5)** Regras positivas que dependem de sema – essas regras são de anotação, como as de (2.2), sendo que seu diferencial é que são mais restritas pois são aplicadas apenas no final e têm como condição uma anotação específica de sema para o termo.

**Exemplo:**

(a regra deve anotar apenas os casos de “campo de Catu” quando “Catu” tiver sema O, deixando os demais casos de lado)

if regex("Catu", token.lemma) and regex("O", token.deps) and regex("de", token.previous\_token.lemma) and regex("campo", token.previous\_token.previous\_token.lemma):

token.deps = "I=CAMPO"

token.previous\_token.deps = "I=CAMPO"

**(2.6)** Regras de limpeza final (acabamentos) – destinam-se a limpar erros facilmente detectáveis pelo formato da anotação IOB2.

**Exemplo:**

# limpar entidade do tipo I quando o termo anterior tem anotação diferente

antes: Santos\_O et\_I al.\_I

depois: Santos\_O et\_O al.\_O

# limpar B=X|I=X B=CAMPO|I=CAMPO (Serra)

Rio de a Serra\_B=CAMPO|I=CAMPO

depois: I=CAMPO

# limpar de|o cujo próximo é O

antes: Bacia\_B de\_I Santos\_I de\_I o\_I Brasil\_O

depois: Bacia\_B de\_I Santos\_I de\_O o\_O Brasil\_O

(Brasil já foi limpado por outra regra anteriormente, restando só o de|o)

# limpar /conj e /flat:name cujo head perdeu sema

antes: Santos\_O e\_O Medina\_B=BACIA/conj, 1986\_O

depois: Santos\_O e\_O Medina\_O, 1986\_O

## **O processo de revisão linguística**

O processo de revisão linguística teve início com a distribuição das entidades com instâncias definidas (BACIA, CAMPO, FLUIDO, FLUIDO DA TERRA, POÇO e UNIDADE CRONOESTRATIGRÁFICA) entre as anotadoras. O primeiro passo foi buscar no corpus – já anotado automaticamente – pela distribuição por lemas do sema desejado. Assim, o objetivo era o de verificar, caso a caso, o que havia sido corretamente anotado com determinada etiqueta.

Nesse processo, também foram filtradas as “colocações” dos semas, isto é, combinações frequentes de palavras associadas a eles. Por exemplo, as buscas pelas colocações da palavra “campo” incluíram sequências como “campo magnético”, “campo de tensões”, ou “trabalho de campo”, que configuram casos em que a palavra “campo” não é entidade. Dependendo da entidade, também foram utilizadas estratégias para a identificação de falsos negativos, como por exemplo a busca por palavras terminadas em “-iano” ou “-oceno” que não haviam recebido a etiqueta UNIDADE CRONOESTRATIGRÁFICA. Frente aos resultados dessas análises, as devidas correções foram realizadas por meio de regras.

Além da anotação automática que passou por um processo de revisão cuidadoso, conforme descrito anteriormente, algumas classes foram inteiramente anotadas por meio de regras criadas manualmente – ou seja, começando pela segunda etapa de anotação, conforme descrito no item a. São elas: ESTRUTURA\_FÍSICA, EVENTO\_PETRO, FLUIDO, FLUIDODATERRA\_i, FLUIDODATERRA\_o, POÇO\_Q, POÇO\_R, POÇO\_T, TEXTURA e TIPO\_POROSIDADE. Em alguns casos, o processo de revisão se manteve o mesmo, ou seja, foram anotadas todas as ocorrências de determinada entidade e, posteriormente, foram revistas aquelas que não deveriam ser anotadas. Em outros casos, a regra de anotação manual foi criada para ser aplicada apenas às ocorrências nas quais a palavra se tratava de fato da entidade.

Durante todo o processo de revisão, os casos de dúvida foram tratados com especialistas da área. Essas consultas aconteceram através da recuperação do contexto de sentenças inteiras e também pela produção de documentos que reuniu, em forma de listas, as palavras que modificam e são modificadas por determinadas entidades. Por exemplo, listas com as palavras modificadas por *petróleo*, *água* e *óleo*:

| **X de petróleo** | **X de água** | **X com óleo** |
| --- | --- | --- |
| Exploração [de petróleo] | Região [de água] | Areia [com óleo] |
| Campo [de petróleo] | Influxo [de água] | Lamparina [com óleo] |
| Inclusão [de petróleo] | Saturação [de água] | Reservatório [com óleo] |
| Expulsão [de petróleo] | Volume [de água] | Zona [com óleo] |

# Resultados

## **a) Novas entidades**

Alimentamos a taxonomia inicial com novas instâncias, usando uma metodologia que combina ontology driven (ou knowledge driven) e data driven.

| **Classes** | **Entidades novas** |
| --- | --- |
| BACIA | 48 |
| CAMPO | 6 |
| ESTRUTURA\_FÍSICA | 3 |
| EVENTO\_PETRO | 1 |
| POÇO | 56 |
| ROCHA | 15 |
| TEXTURA | 4 |
| TIPO\_POROSIDADE | 2 |
| UNIDADE\_CRONO | 100 |
| UNIDADE\_LITO | 64 |

## **b) Distribuição das entidades anotadas no corpus**

| **Classes** | **Entidades anotadas** |
| --- | --- |
| BACIA | 4057 |
| CAMPO | 708 |
| ESTRUTURA\_FÍSICA | 2042 |
| EVENTO\_PETRO | 492 |
| FLUIDO | 175 |
| FLUIDODATERRA\_i | 1374 |
| FLUIDODATERRA\_o | 246 |
| NÃOCONSOLID | 1036 |
| POÇO | 1230 |
| POÇO\_Q | 9 |
| POÇO\_R | 3 |
| POÇO\_T | 35 |
| ROCHA | 2774 |
| TEXTURA | 141 |
| TIPO\_POROSIDADE | 23 |
| UNIDADE\_CRONO | 2920 |
| UNIDADE\_LITO | 1492 |

# Avaliação do anotador por regras

## **Frequência de aplicação das regras**

No total, foram desenvolvidas 1939 regras. A distribuição detalhada das regras por número de vezes que foram aplicadas está no Anexo 1: 1091 regras (56,2%) foram aplicadas apenas uma vez, sendo regras de difícil generalização e que, portanto, podem não se aplicar a outros corpora.

## **Entidades polissêmicas**

Há entidades no corpus que ora estão anotadas como de um tipo, ora de outro tipo. Abaixo encontra-se uma lista das 196 entidades (por lema) com ambiguidade por ordem de frequência no corpus.

bacia: 1327

BACIA - 1322

O - 5

2: 886

O - 885

UNIDADE\_LITO - 1

falha: 662

ESTRUTURA\_FÍSICA - 659

O - 3

poço: 646

POÇO - 644

O - 2

petróleo: 455

FLUIDODATERRA\_i - 339

O - 116

rocha: 449

ROCHA - 448

O - 1

estratigráfico: 442

O - 433

POÇO\_T - 9

água: 441

O - 204

FLUIDODATERRA\_o - 151

FLUIDODATERRA\_i - 83

FLUIDO - 3

óleo: 430

FLUIDODATERRA\_i - 425

O - 5

estrutura: 357

ESTRUTURA\_FÍSICA - 351

O - 6

campo: 332

CAMPO - 201

O - 131

formação: 313

O - 312

UNIDADE\_LITO - 1

Sul: 252

O - 251

BACIA - 1

gás: 213

FLUIDODATERRA\_i - 208

O - 4

FLUIDODATERRA\_o - 1

produção: 180

O - 179

POÇO\_R - 1

geração: 177

EVENTO\_PETRO - 94

O - 83

fluido: 174

FLUIDO - 171

O - 3

Santos: 166

O - 126

BACIA - 40

Alagoas: 159

BACIA - 105

O - 32

UNIDADE\_CRONO - 22

extensão: 154

O - 152

POÇO\_T - 2

fratura: 142

ESTRUTURA\_FÍSICA - 140

O - 2

carbonato: 118

ROCHA - 102

O - 16

migração: 115

EVENTO\_PETRO - 101

O - 14

Salvador: 112

O - 109

UNIDADE\_LITO - 3

Campos: 112

BACIA - 71

O - 41

siliciclástico: 106

O - 94

ROCHA - 12

contato: 103

O - 98

ESTRUTURA\_FÍSICA - 5

gerador: 91

EVENTO\_PETRO - 65

O - 26

Norte: 89

O - 83

BACIA - 6

vulcânico: 87

O - 68

ROCHA - 19

série: 84

O - 83

UNIDADE\_CRONO - 1

evaporito: 83

ROCHA - 82

O - 1

maciço: 78

ESTRUTURA\_FÍSICA - 77

O - 1

Natal: 70

O - 69

BACIA - 1

Potiguar: 68

BACIA - 67

O - 1

seco: 66

O - 62

POÇO\_Q - 4

produtor: 66

O - 64

POÇO\_R - 2

Paraná: 66

O - 54

BACIA - 12

Cabo: 64

O - 63

BACIA - 1

preservação: 63

O - 42

EVENTO\_PETRO - 21

Sergipe: 62

BACIA - 36

O - 26

Recôncavo: 62

BACIA - 43

O - 19

Santoniano: 61

UNIDADE\_CRONO - 60

O - 1

especial: 58

O - 57

POÇO\_T - 1

Pernambuco: 58

O - 57

BACIA - 1

Sergi: 56

O - 52

UNIDADE\_LITO - 4

Amazonas: 50

O - 35

BACIA - 15

Maastrichtiano: 50

UNIDADE\_CRONO - 49

O - 1

Arco: 48

O - 47

ESTRUTURA\_FÍSICA - 1

Turoniano: 47

UNIDADE\_CRONO - 46

O - 1

Espírito Santo: 45

BACIA - 44

UNIDADE\_LITO - 1

Terciário: 44

UNIDADE\_CRONO - 41

O - 3

alongar: 42

O - 40

TEXTURA - 2

Paraíba: 41

O - 38

BACIA - 3

Pelotas: 41

BACIA - 29

O - 12

Tucano: 40

BACIA - 38

O - 2

Almada: 40

O - 26

BACIA - 14

Curitiba: 39

O - 38

BACIA - 1

Bacias: 38

O - 30

BACIA - 8

Ceará: 35

O - 23

BACIA - 12

pioneiro: 33

POÇO\_T - 17

O - 16

estratificação: 31

ESTRUTURA\_FÍSICA - 25

O - 6

terrígeno: 29

O - 18

ROCHA - 11

Cenozóico: 29

UNIDADE\_CRONO - 28

O - 1

Cenomaniano: 26

UNIDADE\_CRONO - 25

O - 1

Ubarana: 26

O - 23

CAMPO - 2

UNIDADE\_LITO - 1

Jatobá: 26

BACIA - 22

O - 4

bioturbação: 25

ESTRUTURA\_FÍSICA - 24

O - 1

well: 24

O - 23

POÇO - 1

Sergipe-Alagoas: 24

BACIA - 23

O - 1

Araripe: 24

O - 13

BACIA - 11

Maranhão: 24

O - 12

BACIA - 12

PP-11: 24

POÇO - 23

O - 1

membro: 22

O - 20

UNIDADE\_LITO - 2

Oligoceno: 22

UNIDADE\_CRONO - 21

O - 1

Bacia: 21

BACIA - 18

O - 3

Mucuri: 20

BACIA - 16

O - 4

Pará: 20

O - 17

BACIA - 3

ígneo: 20

O - 18

ROCHA - 2

Central: 19

O - 14

BACIA - 5

Namorado: 19

CAMPO - 17

O - 2

Médio: 17

O - 15

UNIDADE\_CRONO - 2

laminação: 17

ESTRUTURA\_FÍSICA - 16

O - 1

água de o mar: 16

FLUIDODATERRA\_o - 14

FLUIDODATERRA\_i - 2

Verde: 16

O - 15

UNIDADE\_LITO - 1

caótico: 16

O - 8

TEXTURA - 8

mosaico: 15

O - 14

TEXTURA - 1

Aratu: 15

UNIDADE\_CRONO - 11

O - 3

CAMPO - 1

Jiquiá: 15

UNIDADE\_CRONO - 14

CAMPO - 1

Irati: 15

O - 10

UNIDADE\_LITO - 5

Buracica: 15

UNIDADE\_CRONO - 9

CAMPO - 5

O - 1

Superior: 14

O - 13

UNIDADE\_CRONO - 1

lenticular: 14

ESTRUTURA\_FÍSICA - 13

O - 1

Roncador: 14

CAMPO - 13

O - 1

Carmópolis: 14

O - 10

CAMPO - 3

UNIDADE\_LITO - 1

explotação: 13

O - 12

POÇO\_T - 1

Curimã: 13

CAMPO - 11

O - 2

Camboriú: 13

O - 12

UNIDADE\_LITO - 1

Recôncavo-Tucano-Jatobá: 13

BACIA - 12

O - 1

Solimões: 12

BACIA - 10

O - 2

Gabão: 12

BACIA - 6

O - 6

intersticial: 11

O - 8

TEXTURA - 3

Formação: 11

O - 9

UNIDADE\_LITO - 2

Ibura: 11

O - 10

UNIDADE\_LITO - 1

Juruá: 11

CAMPO - 10

O - 1

Amazônia: 11

O - 10

BACIA - 1

H2S: 11

O - 8

FLUIDODATERRA\_o - 3

Congo: 10

BACIA - 5

O - 5

Maceió: 10

O - 8

UNIDADE\_LITO - 2

Mesoalbiano: 10

UNIDADE\_CRONO - 8

O - 2

Camaçari: 10

O - 9

CAMPO - 1

Bonito: 10

O - 8

CAMPO - 2

CH4: 10

O - 9

FLUIDODATERRA\_o - 1

molde: 9

ESTRUTURA\_FÍSICA - 5

O - 4

Dom João: 9

CAMPO - 5

UNIDADE\_CRONO - 4

Jacuípe: 9

BACIA - 6

O - 3

Triunfo: 9

O - 8

UNIDADE\_LITO - 1

Riachuelo: 8

O - 5

CAMPO - 2

UNIDADE\_LITO - 1

Muribeca: 8

O - 7

UNIDADE\_LITO - 1

N2: 8

O - 5

FLUIDODATERRA\_o - 3

Campo: 8

O - 5

CAMPO - 3

Badejo: 8

CAMPO - 4

O - 4

Garoupa: 8

CAMPO - 6

O - 2

Regência: 7

O - 5

UNIDADE\_LITO - 2

Penedo: 7

UNIDADE\_LITO - 5

O - 2

Valanginiano: 7

UNIDADE\_CRONO - 6

O - 1

Maestrichtiano: 7

UNIDADE\_CRONO - 6

O - 1

Itamaracá: 7

O - 6

UNIDADE\_LITO - 1

Pilar: 7

O - 4

CAMPO - 3

Well: 6

O - 4

POÇO - 2

Alagamar: 6

UNIDADE\_LITO - 4

O - 2

Barreiras: 6

UNIDADE\_LITO - 5

O - 1

contração: 6

O - 5

TIPO\_POROSIDADE - 1

Atokan: 6

UNIDADE\_CRONO - 5

O - 1

Andar: 6

O - 5

UNIDADE\_CRONO - 1

Illinois: 6

O - 5

BACIA - 1

Piauí: 6

BACIA - 3

O - 3

Rocha: 6

O - 5

ROCHA - 1

Pirambóia: 6

UNIDADE\_LITO - 3

O - 3

Poço: 6

O - 5

POÇO - 1

Taciba: 6

O - 3

UNIDADE\_LITO - 3

Paracuru: 5

UNIDADE\_LITO - 4

O - 1

Itaituba: 5

UNIDADE\_LITO - 4

O - 1

Guamaré: 5

UNIDADE\_LITO - 2

O - 2

CAMPO - 1

HL-1: 5

POÇO - 4

O - 1

Botucatu: 5

UNIDADE\_LITO - 3

O - 2

Caravelas: 5

O - 3

UNIDADE\_LITO - 2

H2: 5

O - 4

FLUIDODATERRA\_o - 1

Taubaté: 5

O - 4

BACIA - 1

Estreito: 5

O - 3

CAMPO - 2

Andirá: 5

UNIDADE\_LITO - 4

O - 1

Eoturoniano: 5

UNIDADE\_CRONO - 4

O - 1

Sydney: 4

O - 3

BACIA - 1

Ilhas: 4

UNIDADE\_LITO - 2

O - 2

Teresina: 4

UNIDADE\_LITO - 2

O - 2

Bananeiras: 4

UNIDADE\_LITO - 3

O - 1

Coqueiro Seco: 4

UNIDADE\_LITO - 3

CAMPO - 1

GTP-8: 4

O - 3

POÇO - 1

Candeias: 4

CAMPO - 3

O - 1

Atum: 4

CAMPO - 3

O - 1

Palmeira: 4

O - 3

CAMPO - 1

Mundaú: 4

O - 2

UNIDADE\_LITO - 1

BACIA - 1

Urucu: 4

O - 3

CAMPO - 1

descobridor: 4

POÇO\_Q - 3

O - 1

meso: 3

O - 2

TEXTURA - 1

Açu: 3

UNIDADE\_LITO - 2

O - 1

dilatação: 3

ESTRUTURA\_FÍSICA - 2

O - 1

água doce: 3

FLUIDODATERRA\_o - 2

FLUIDODATERRA\_i - 1

Pendência: 3

O - 2

UNIDADE\_LITO - 1

água salgado: 3

FLUIDODATERRA\_i - 2

FLUIDODATERRA\_o - 1

P3: 3

POÇO - 2

O - 1

Cenozóicos: 3

O - 2

UNIDADE\_CRONO - 1

Redondo: 3

O - 2

UNIDADE\_LITO - 1

Namíbia: 3

O - 2

BACIA - 1

Urucutuca: 3

O - 2

UNIDADE\_LITO - 1

Barreirinha: 3

O - 2

BACIA - 1

Enchova: 3

CAMPO - 2

O - 1

Mafra: 3

UNIDADE\_LITO - 2

O - 1

Paraguaçu: 3

UNIDADE\_LITO - 2

O - 1

Palermo: 3

UNIDADE\_LITO - 2

O - 1

Canguaretama: 3

BACIA - 2

O - 1

Cedro: 3

BACIA - 2

O - 1

sub-comercial: 3

O - 2

POÇO\_Q - 1

Lapa: 3

O - 2

CAMPO - 1

pull-apart: 2

O - 1

ESTRUTURA\_FÍSICA - 1

alveolar: 2

TEXTURA - 1

O - 1

São Mateus: 2

CAMPO - 1

UNIDADE\_LITO - 1

Tibau: 2

UNIDADE\_LITO - 1

O - 1

Furnas: 2

O - 1

UNIDADE\_LITO - 1

N’Dombo: 2

UNIDADE\_LITO - 1

O - 1

Batinga: 2

UNIDADE\_LITO - 1

O - 1

Lutetiano: 2

UNIDADE\_CRONO - 1

O - 1

Kunguriano: 2

UNIDADE\_CRONO - 1

O - 1

Afligidos: 2

UNIDADE\_LITO - 1

O - 1

Calumbi: 2

UNIDADE\_LITO - 1

O - 1

Pré-Cambriana: 2

O - 1

UNIDADE\_CRONO - 1

# Referências

CAVALCANTI, Tatiana; SILVEIRA, Aline; DE SOUZA, Elvis; FREITAS, Cláudia. Os limites da palavra e da sentença no processamento automático de textos. Revista Brasileira de Iniciação Científica, [S. l.], v. 8, p. e021033, 2021. Disponível em: https://periodicoscientificos.itp.ifsp.edu.br/index.php/rbic/article/view/348. Acesso em: 15 out. 2021.

DE SOUZA, Elvis; FREITAS, Cláudia. Polishing the gold – how much revision do we need in treebanks?. In: Procedings of the Universal Dependencies Brazilian Festival, p. 1–11, Fortaleza, Brazil. Association for Computational Linguistics, 2022. Disponível em: https://aclanthology.org/2022.udfestbr-1.2/. Acesso em: 11 de jun. 2022.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton and Christopher D. Manning. 2020. [Stanza: A Python Natural Language Processing Toolkit for Many Human Languages.](https://arxiv.org/abs/2003.07082) In Association for Computational Linguistics (ACL) System Demonstrations. 2020. [[pdf](https://nlp.stanford.edu/pubs/qi2020stanza.pdf)][[bib](https://nlp.stanford.edu/pubs/qi2020stanza.bib)]

# Anexo 1

**regra (linha), frequência**

2929, 2701

5867, 1069

2899, 978

627, 970

2305, 870

2873, 773

2405, 757

2872, 746

2853, 718

3007, 696

222, 670

4752, 660

2911, 614

1883, 592

761, 576

2894, 540

2892, 530

2893, 530

2847, 516

2510, 431

1672, 413

2322, 361

1142, 347

2316, 236

767, 232

4769, 230

2501, 225

4048, 218

2864, 209

2866, 209

2188, 206

3977, 205

3297, 197

5878, 192

1863, 178

526, 174

2649, 171

335, 167

336, 167

4954, 158

220, 156

3600, 149

423, 145

424, 145

2300, 142

2665, 141

216, 139

5888, 135

4943, 132

765, 130

426, 123

427, 123

5712, 120

108, 118

2058, 116

763, 108

2377, 107

2883, 106

2884, 106

2886, 106

5883, 101

755, 96

2970, 95

396, 91

655, 90

3814, 88

3399, 88

3236, 85

3237, 85

783, 82

3462, 81

3693, 80

385, 78

3089, 72

2657, 70

2658, 70

2596, 68

3982, 68

2982, 66

2983, 66

5311, 66

2395, 60

2372, 57

785, 55

3359, 54

3980, 54

3292, 52

2425, 51

2400, 51

2978, 50

2979, 50

4946, 49

2919, 48

2921, 48

5099, 48

809, 47

810, 47

4028, 46

571, 45

575, 45

576, 45

799, 41

4801, 41

3431, 41

565, 40

4820, 39

3681, 38

4825, 37

5611, 37

2587, 37

5473, 36

4774, 35

771, 35

2678, 35

3442, 34

5873, 33

2965, 32

331, 32

2330, 32

2108, 30

2109, 30

2191, 30

3782, 30

2319, 29

4023, 29

5119, 28

2038, 27

3043, 26

3383, 26

2297, 26

4504, 26

3414, 25

5067, 25

757, 25

3994, 24

863, 24

864, 24

773, 24

2390, 23

567, 23

4763, 23

1685, 22

1885, 22

1886, 22

803, 21

804, 21

891, 20

892, 20

801, 20

2768, 20

2769, 20

3501, 20

1009, 19

2529, 19

2530, 19

5510, 19

3285, 18

3286, 18

3287, 18

1681, 18

1682, 18

1683, 18

1373, 18

1374, 18

1375, 18

1376, 18

1377, 18

1378, 18

3183, 18

2739, 18

2740, 18

5068, 17

1381, 17

1382, 17

1383, 17

1384, 17

1385, 17

1386, 17

1618, 17

1471, 16

1472, 16

1473, 16

1474, 16

1475, 16

1476, 16

234, 16

5591, 16

5592, 16

775, 16

3429, 16

3321, 16

569, 16

3341, 15

1249, 15

3174, 15

1867, 15

2342, 15

3553, 15

3554, 15

3555, 15

3556, 15

1877, 15

4176, 15

2545, 14

2546, 14

2547, 14

2548, 14

2759, 14

2760, 14

1879, 14

5573, 14

5660, 14

2382, 14

506, 14

507, 14

508, 14

509, 14

3884, 14

3885, 14

859, 13

860, 13

3868, 13

3869, 13

729, 13

3671, 13

3065, 13

3098, 13

3973, 13

3514, 13

5763, 13

5764, 13

4754, 13

1600, 12

3083, 12

5461, 12

3113, 12

5106, 12

2461, 12

979, 12

980, 12

1974, 12

1630, 12

3361, 12

2361, 12

2339, 12

4843, 12

3403, 12

1270, 11

2041, 11

5530, 11

4030, 11

2194, 11

2195, 11

3017, 11

3015, 11

3689, 10

3178, 10

3614, 10

4367, 10

4967, 10

4413, 10

1951, 10

2354, 10

3369, 10

4452, 10

1977, 9

5832, 9

2985, 9

2986, 9

2493, 9

2254, 9

608, 9

3801, 9

5617, 9

438, 9

4624, 9

4625, 9

3067, 9

3774, 9

4803, 9

4507, 9

3023, 8

3111, 8

1012, 8

3922, 8

4553, 8

359, 8

1967, 8

1917, 8

3777, 8

777, 8

524, 8

4650, 8

4477, 8

3367, 8

4453, 8

3975, 8

405, 7

3769, 7

4510, 7

2525, 7

2526, 7

2121, 7

3150, 7

1389, 7

1391, 7

3391, 7

3748, 7

3543, 7

4630, 7

4644, 7

1916, 7

2385, 7

4347, 7

4348, 7

4349, 7

473, 7

474, 7

4934, 7

4935, 7

4936, 7

4937, 7

3761, 7

3687, 7

4209, 7

5132, 7

5133, 7

3158, 7

3568, 7

3569, 7

3570, 7

3571, 7

2653, 7

2654, 7

5566, 7

5568, 7

3809, 6

5064, 6

3127, 6

2127, 6

3722, 6

1913, 6

1914, 6

3063, 6

3738, 6

3247, 6

3248, 6

1441, 6

1442, 6

1443, 6

5439, 6

4065, 6

4066, 6

1479, 6

1480, 6

1481, 6

1482, 6

1483, 6

1484, 6

3074, 6

3323, 6

2355, 6

4837, 6

2782, 6

2783, 6

2784, 6

2785, 6

2786, 6

3125, 6

2274, 6

781, 6

2554, 6

3896, 6

3897, 6

4139, 6

4841, 6

2629, 5

4577, 5

4578, 5

1320, 5

4931, 5

3860, 5

3861, 5

3444, 5

4764, 5

5708, 5

5709, 5

1252, 5

1253, 5

2267, 5

4474, 5

2989, 5

4003, 5

4004, 5

4005, 5

2357, 5

2358, 5

5758, 5

1464, 5

1465, 5

1466, 5

1467, 5

1468, 5

5697, 5

3021, 5

3912, 5

3673, 5

5183, 5

5184, 5

3076, 5

3134, 5

469, 5

4891, 5

3048, 5

1494, 5

1495, 5

1496, 5

1497, 5

1498, 5

1394, 5

1395, 5

1396, 5

1397, 5

1398, 5

1399, 5

536, 5

2505, 5

2661, 5

2662, 5

348, 5

349, 5

1869, 5

1870, 5

4822, 5

4823, 5

5393, 4

5396, 4

1021, 4

2048, 4

5202, 4

3961, 4

2432, 4

1027, 4

2175, 4

2176, 4

2177, 4

2178, 4

2179, 4

1030, 4

4097, 4

3659, 4

3663, 4

3907, 4

1326, 4

2988, 4

4253, 4

792, 4

793, 4

794, 4

1562, 4

1563, 4

1564, 4

1565, 4

1566, 4

1567, 4

2333, 4

5125, 4

586, 4

587, 4

3277, 4

251, 4

252, 4

3880, 4

3881, 4

4791, 4

4792, 4

4793, 4

2081, 4

3753, 4

3954, 4

2308, 4

2309, 4

2021, 4

2022, 4

753, 4

4331, 4

4332, 4

4333, 4

4334, 4

1254, 4

5044, 4

5151, 4

5152, 4

5153, 4

492, 4

3642, 4

3643, 4

3644, 4

1624, 4

1944, 4

1945, 4

1960, 4

1961, 4

1962, 4

1963, 4

3318, 4

5576, 4

4647, 4

2198, 4

2199, 4

2200, 4

1919, 4

1615, 3

1669, 3

2620, 3

214, 3

5827, 3

622, 3

623, 3

2449, 3

2450, 3

2451, 3

230, 3

2087, 3

559, 3

3712, 3

1390, 3

3120, 3

2101, 3

1891, 3

3848, 3

3849, 3

3050, 3

2518, 3

2519, 3

2013, 3

4342, 3

4343, 3

4344, 3

967, 3

968, 3

182, 3

3876, 3

3877, 3

2602, 3

955, 3

956, 3

5017, 3

5018, 3

2351, 3

2345, 3

3451, 3

1284, 3

3460, 3

2280, 3

2034, 3

2035, 3

5320, 3

5321, 3

5322, 3

5323, 3

1355, 3

1356, 3

1357, 3

1358, 3

1359, 3

5840, 3

1603, 3

779, 3

556, 3

557, 3

4361, 3

5281, 3

5282, 3

5283, 3

5284, 3

5285, 3

3085, 3

2270, 3

2271, 3

2078, 3

4358, 3

4961, 3

2419, 3

2028, 3

2029, 3

2030, 3

2031, 3

2025, 3

4748, 3

2114, 3

2366, 3

5718, 3

5719, 3

5720, 3

5721, 3

4845, 3

3176, 3

2248, 3

2249, 3

2250, 3

2413, 3

5426, 3

3485, 3

4215, 3

2416, 3

2539, 3

3577, 3

459, 3

4171, 3

4172, 3

3624, 3

4559, 3

1329, 3

5779, 3

2640, 3

2244, 3

2245, 3

1299, 3

398, 3

399, 3

3796, 3

4809, 3

2003, 2

5061, 2

1633, 2

4839, 2

5890, 2

4456, 2

3832, 2

3833, 2

3852, 2

3853, 2

2581, 2

4163, 2

2124, 2

2055, 2

5464, 2

3717, 2

2701, 2

2702, 2

4636, 2

3628, 2

3902, 2

228, 2

4403, 2

4404, 2

4405, 2

5005, 2

3941, 2

4616, 2

1120, 2

1121, 2

1122, 2

1033, 2

4153, 2

5705, 2

3052, 2

4897, 2

5745, 2

5746, 2

2148, 2

927, 2

928, 2

2728, 2

867, 2

868, 2

4381, 2

4382, 2

4383, 2

4384, 2

4978, 2

4979, 2

4980, 2

4981, 2

999, 2

1000, 2

3433, 2

4364, 2

4970, 2

4971, 2

4849, 2

4850, 2

4851, 2

3785, 2

5399, 2

4396, 2

5002, 2

5319, 2

3033, 2

5071, 2

5072, 2

5073, 2

5265, 2

5266, 2

5267, 2

5268, 2

5269, 2

3508, 2

1276, 2

3788, 2

3061, 2

4723, 2

4724, 2

2967, 2

1102, 2

1103, 2

3059, 2

4387, 2

4388, 2

4389, 2

4390, 2

4964, 2

1995, 2

1996, 2

1997, 2

1998, 2

1999, 2

2285, 2

2617, 2

3806, 2

4250, 2

3630, 2

3148, 2

3149, 2

5194, 2

1923, 2

1924, 2

3188, 2

2939, 2

3840, 2

3841, 2

3265, 2

3266, 2

4421, 2

4422, 2

1666, 2

4487, 2

939, 2

940, 2

2790, 2

2791, 2

2792, 2

2793, 2

1992, 2

4879, 2

4880, 2

4881, 2

4882, 2

4940, 2

2061, 2

2062, 2

3936, 2

5715, 2

631, 2

3312, 2

3313, 2

2732, 2

3864, 2

3865, 2

1411, 2

1412, 2

1413, 2

1414, 2

1415, 2

1416, 2

4592, 2

1573, 2

1574, 2

1575, 2

4167, 2

3139, 2

2130, 2

5645, 2

5639, 2

2006, 2

2972, 2

2065, 2

4258, 2

3758, 2

2443, 2

2444, 2

2445, 2

2446, 2

5423, 2

1585, 2

4032, 2

4501, 2

5470, 2

3377, 2

3416, 2

4581, 2

665, 2

353, 2

4328, 2

2749, 2

2750, 2

1872, 2

4034, 2

4035, 2

4196, 2

169, 2

3205, 2

3453, 2

3395, 2

5477, 2

5478, 2

5402, 2

4543, 2

5731, 2

5081, 2

5082, 2

5083, 2

5084, 2

1660, 2

2708, 2

2229, 2

2230, 2

2231, 2

2232, 2

2233, 2

5295, 2

1902, 2

1903, 2

1910, 2

1911, 2

5230, 2

5231, 2

4833, 2

3636, 2

3638, 2

5767, 2

1058, 2

1059, 2

1314, 2

2652, 2

2236, 2

2237, 2

2238, 2

2239, 2

2240, 2

2241, 2

5291, 2

5292, 2

538, 2

539, 2

540, 2

541, 2

4323, 1

3683, 1

4181, 1

387, 1

388, 1

3984, 1

871, 1

872, 1

1024, 1

5198, 1

5491, 1

267, 1

269, 1

271, 1

272, 1

273, 1

274, 1

5550, 1

987, 1

988, 1

436, 1

254, 1

256, 1

258, 1

259, 1

260, 1

261, 1

263, 1

1362, 1

1363, 1

1364, 1

1365, 1

1366, 1

1367, 1

1368, 1

1369, 1

1370, 1

4114, 1

4115, 1

1339, 1

1340, 1

1341, 1

3389, 1

3193, 1

1716, 1

1719, 1

1721, 1

1722, 1

1723, 1

1724, 1

1725, 1

5481, 1

5482, 1

5483, 1

5484, 1

5027, 1

5028, 1

5372, 1

5373, 1

3766, 1

3198, 1

3956, 1

3200, 1

2097, 1

2093, 1

2094, 1

1055, 1

4716, 1

4717, 1

2090, 1

4685, 1

3527, 1

236, 1

2075, 1

3743, 1

3057, 1

1045, 1

4039, 1

959, 1

960, 1

2705, 1

5272, 1

2496, 1

2477, 1

2258, 1

2259, 1

519, 1

2104, 1

2105, 1

2100, 1

2289, 1

2290, 1

3844, 1

3845, 1

1407, 1

1408, 1

3483, 1

4132, 1

4133, 1

5595, 1

5596, 1

5597, 1

3118, 1

1956, 1

1957, 1

1958, 1

2154, 1

1311, 1

601, 1

4900, 1

5215, 1

5216, 1

4408, 1

4409, 1

4410, 1

1696, 1

1699, 1

1701, 1

1702, 1

1706, 1

1709, 1

1712, 1

5579, 1

5580, 1

5581, 1

3646, 1

3647, 1

3648, 1

3649, 1

3650, 1

3651, 1

3653, 1

3654, 1

3656, 1

3657, 1

3109, 1

372, 1

825, 1

827, 1

829, 1

4857, 1

4858, 1

5010, 1

5011, 1

5147, 1

5148, 1

5494, 1

975, 1

976, 1

5536, 1

5626, 1

5627, 1

5503, 1

5504, 1

4284, 1

4337, 1

4338, 1

3091, 1

4298, 1

813, 1

4110, 1

4111, 1

4290, 1

4731, 1

4732, 1

5858, 1

5859, 1

4532, 1

4533, 1

4534, 1

2683, 1

2684, 1

2590, 1

736, 1

737, 1

464, 1

465, 1

1305, 1

2165, 1

2161, 1

2162, 1

2144, 1

2145, 1

895, 1

896, 1

4313, 1

2723, 1

2724, 1

2725, 1

2719, 1

2720, 1

2729, 1

2584, 1

3566, 1

3872, 1

3873, 1

3986, 1

3987, 1

3988, 1

3989, 1

3990, 1

142, 1

919, 1

920, 1

963, 1

964, 1

935, 1

936, 1

915, 1

916, 1

923, 1

924, 1

991, 1

992, 1

931, 1

932, 1

911, 1

912, 1

971, 1

972, 1

879, 1

880, 1

875, 1

876, 1

5316, 1

2689, 1

883, 1

884, 1

903, 1

904, 1

185, 1

187, 1

206, 1

2560, 1

2561, 1

3856, 1

3857, 1

2764, 1

2765, 1

146, 1

5507, 1

3707, 1

1544, 1

1545, 1

1548, 1

4663, 1

1551, 1

1557, 1

1558, 1

1559, 1

1534, 1

1540, 1

1541, 1

2542, 1

4416, 1

4417, 1

4418, 1

2614, 1

1048, 1

5224, 1

853, 1

854, 1

855, 1

856, 1

1795, 1

1797, 1

1799, 1

1074, 1

1075, 1

1287, 1

1881, 1

1184, 1

1187, 1

1190, 1

1191, 1

4865, 1

5513, 1

4287, 1

4925, 1

983, 1

984, 1

5109, 1

5110, 1

2224, 1

2225, 1

2226, 1

2515, 1

4354, 1

4355, 1

4957, 1

4958, 1

2623, 1

2564, 1

4619, 1

3479, 1

2157, 1

2158, 1

2151, 1

210, 1

3039, 1

1302, 1

1815, 1

1817, 1

1151, 1

1152, 1

1153, 1

1578, 1

1579, 1

1580, 1

1581, 1

1582, 1

2551, 1

1594, 1

499, 1

4705, 1

4706, 1

1082, 1

1083, 1

4441, 1

4228, 1

1039, 1

4274, 1

5807, 1

3096, 1

4266, 1

218, 1

3170, 1

4279, 1

4126, 1

4127, 1

4128, 1

4129, 1

4906, 1

3214, 1

5685, 1

5174, 1

5175, 1

5681, 1

5682, 1

5338, 1

1018, 1

1098, 1

1099, 1

1015, 1

232, 1

3632, 1

3633, 1

3634, 1

951, 1

952, 1

1042, 1

5257, 1

5258, 1

1970, 1

1971, 1

887, 1

888, 1

1663, 1

5749, 1

5750, 1

5751, 1

4301, 1

4302, 1

3019, 1

5113, 1

4370, 1

4371, 1

4974, 1

4975, 1

4991, 1

4992, 1

4993, 1

4994, 1

4984, 1

4985, 1

4986, 1

4987, 1

4988, 1

4885, 1

1051, 1

1052, 1

3078, 1

5614, 1

1425, 1

1426, 1

1427, 1

5167, 1

5168, 1

1501, 1

1502, 1

1503, 1

1504, 1

1505, 1

1508, 1

1509, 1

1511, 1

1512, 1

1513, 1

1514, 1

1515, 1

2209, 1

2210, 1

2211, 1

3535, 1

1804, 1

1806, 1

1808, 1

1809, 1

1811, 1

3028, 1

5525, 1

1279, 1

1131, 1

1135, 1

5355, 1

822, 1

1894, 1

1895, 1

1896, 1

1897, 1

3410, 1

3411, 1

3412, 1

5849, 1

4922, 1

2626, 1

2522, 1

2593, 1

3930, 1

4223, 1

1090, 1

1091, 1

5138, 1

3562, 1

4201, 1

2294, 1

2635, 1

2203, 1

2204, 1

2205, 1

2206, 1

2576, 1

2084, 1

2573, 1

2567, 1

2605, 1

2570, 1

5519, 1

1266, 1

1267, 1

4178, 1

3407, 1

3408, 1

3669, 1

5103, 1

2348, 1

212, 1

3551, 1

4691, 1

4692, 1

4693, 1

4694, 1

4695, 1

4293, 1

4220, 1

194, 1

198, 1

4600, 1

5386, 1

5387, 1

5724, 1

5725, 1

5726, 1

484, 1

485, 1

486, 1

1989, 1

4231, 1

3172, 1

4014, 1

4245, 1

1689, 1

1692, 1

4913, 1

4869, 1

4556, 1

5244, 1

4835, 1

4653, 1

1263, 1

2557, 1

3102, 1

2312, 1

2313, 1

3926, 1

4399, 1

4400, 1

3928, 1

5604, 1

4607, 1

4608, 1

4903, 1

689, 1

2536, 1

4165, 1

2219, 1

2220, 1

2221, 1

2016, 1

2017, 1

2018, 1

2733, 1

4105, 1

1848, 1

1849, 1

1850, 1

1851, 1

1852, 1

1853, 1

1854, 1

1855, 1

1856, 1

1857, 1

1858, 1

731, 1

1487, 1

1488, 1

1489, 1

1490, 1

1491, 1

2262, 1

2263, 1

2264, 1

2214, 1

2215, 1

2440, 1

2422, 1

3503, 1

3539, 1

1675, 1

4701, 1

4702, 1

1317, 1

3242, 1

3435, 1

3436, 1

3437, 1

3438, 1

947, 1

948, 1

2002, 1

1627, 1

618, 1

1323, 1

3363, 1

3730, 1

543, 1

544, 1

545, 1

3997, 1

5487, 1

5488, 1

1086, 1

1087, 1

4307, 1

4308, 1

4750, 1

5159, 1

1642, 1

2133, 1

2435, 1

1740, 1

1743, 1

2010, 1

5642, 1

1754, 1

1757, 1

5651, 1

5652, 1

5653, 1

5654, 1

5655, 1

5622, 1

5623, 1

1066, 1

1067, 1

5053, 1

642, 1

643, 1

1767, 1

1769, 1

177, 1

1773, 1

2044, 1

2045, 1

1747, 1

1750, 1

2776, 1

2777, 1

2778, 1

2692, 1

2693, 1

2694, 1

2695, 1

2696, 1

2697, 1

2698, 1

5739, 1

5740, 1

3958, 1

4597, 1

5171, 1

393, 1

394, 1

2007, 1

2711, 1

2712, 1

2713, 1

2714, 1

2715, 1

2716, 1

907, 1

908, 1

4666, 1

4562, 1

4563, 1

3132, 1

292, 1

294, 1

296, 1

297, 1

298, 1

299, 1

300, 1

302, 1

2068, 1

3888, 1

3889, 1

1609, 1

4261, 1

4449, 1

5187, 1

4894, 1

995, 1

996, 1

1006, 1

208, 1

5634, 1

5635, 1

5636, 1

4526, 1

4527, 1

4529, 1

4522, 1

4523, 1

842, 1

843, 1

844, 1

846, 1

847, 1

4515, 1

3924, 1

1402, 1

1403, 1

1404, 1

4007, 1

4008, 1

4009, 1

4010, 1

4011, 1

2941, 1

3549, 1

5431, 1

5435, 1

3541, 1

2480, 1

3575, 1

3576, 1

3578, 1

3533, 1

4026, 1

4498, 1

4490, 1

3589, 1

3590, 1

5141, 1

5142, 1

4861, 1

4862, 1

2483, 1

2484, 1

2485, 1

2486, 1

2467, 1

2468, 1

2469, 1

2470, 1

2454, 1

2455, 1

3104, 1

2533, 1

4495, 1

3537, 1

5162, 1

5350, 1

5352, 1

5360, 1

769, 1

3209, 1

5363, 1

5364, 1

5365, 1

5366, 1

4817, 1

4818, 1

3328, 1

429, 1

5812, 1

2632, 1

4234, 1

5558, 1

4121, 1

4122, 1

4123, 1

5754, 1

5755, 1

5648, 1

1257, 1

1258, 1

1259, 1

1260, 1

3560, 1

1606, 1

374, 1

515, 1

2117, 1

2118, 1

1591, 1

3337, 1

3339, 1

1117, 1

249, 1

3035, 1

2168, 1

1940, 1

1941, 1

1942, 1

5607, 1

5608, 1

4566, 1

4567, 1

4568, 1

4709, 1

3679, 1

4050, 1

2138, 1

2139, 1

2140, 1

2141, 1

241, 1

4919, 1

5600, 1

5601, 1

200, 1

204, 1

2608, 1

1570, 1

3564, 1

1522, 1

3665, 1

4041, 1

4193, 1

4744, 1

4745, 1

4746, 1

2473, 1

2474, 1

2489, 1

2490, 1

2182, 1

2183, 1

2184, 1

2185, 1

3529, 1

1636, 1

1947, 1

1308, 1

1296, 1

1293, 1

3606, 1

3626, 1

3627, 1

816, 1

1201, 1

1202, 1

1203, 1

1204, 1

1205, 1

1206, 1

1207, 1

1208, 1

1225, 1

1226, 1

1227, 1

1213, 1

1216, 1

1219, 1

1221, 1

238, 1

4318, 1

4603, 1

4174, 1

1078, 1

1079, 1

3819, 1

4151, 1

3343, 1

3344, 1

3345, 1

4712, 1

4713, 1

3401, 1

5497, 1

5500, 1

747, 1

1094, 1

1095, 1

4462, 1

1345, 1

1346, 1

1347, 1

1348, 1

1349, 1

1350, 1

819, 1

1177, 1

1178, 1

1179, 1

1180, 1

1181, 1

5190, 1

5191, 1

450, 1

461, 1

462, 1

4584, 1

190, 1

4611, 1

3152, 1

3153, 1

1678, 1

2738, 1

5561, 1

1821, 1

1822, 1

1823, 1

1824, 1

1826, 1

3836, 1

3837, 1

3828, 1

3829, 1

3824, 1

3825, 1

1657, 1

4587, 1

4571, 1

4572, 1

4573, 1

4574, 1

4688, 1

2991, 1

2992, 1

4433, 1

4435, 1

4436, 1

4437, 1

4101, 1

1651, 1

1652, 1

1653, 1

1654, 1

2744, 1

2745, 1

4546, 1

4547, 1

4548, 1

4549, 1

5076, 1

5077, 1

5078, 1

5093, 1

5087, 1

5088, 1

5089, 1

5090, 1

3892, 1

3893, 1

4679, 1

4680, 1

1730, 1

1733, 1

1736, 1

1588, 1

1639, 1

1273, 1

5690, 1

5219, 1

5220, 1

4727, 1

4728, 1

3531, 1

899, 1

900, 1

1036, 1

4660, 1

4446, 1

1621, 1

3920, 1

1290, 1

1836, 1

1837, 1

1839, 1

1840, 1

1841, 1

1842, 1

1843, 1

1844, 1

605, 1

606, 1

2959, 1

1905, 1

1906, 1

1907, 1

1908, 1

1597, 1

5800, 1

5801, 1

5784, 1

5785, 1

3793, 1

245, 1

4459, 1

5442, 1

5670, 1

5822, 1

5671, 1

4043, 1

4044, 1

148, 1

149, 1

2797, 1

2798, 1

2799, 1

4242, 1

4271, 1

2369, 1

2945, 1

5542, 1

162, 1

5788, 1

5789, 1

5790, 1

5791, 1

5792, 1

5795, 1

5796, 1

5797, 1

1777, 1

1779, 1

1781, 1

1782, 1

1786, 1

1788, 1

1790, 1

1791, 1

4206, 1

4639, 1

2336, 1

4212, 1

4847, 1

3702, 1

3675, 1

4656, 1

5817, 1

247, 1

243, 1

4046, 1

2408, 1

833, 1

835, 1

837, 1

838, 1

839, 1

1760, 1

1761, 1

1234, 1

1235, 1

1236, 1

1238, 1

1240, 1

4020, 1

159, 1

1931, 1

1932, 1

1933, 1

1934, 1

1935, 1

1936, 1

1937, 1

1645, 1

5674, 1

5675, 1

1106, 1

1108, 1

1111, 1

1114, 1

4537, 1

4239, 1

943, 1

944, 1

4161, 1

1128, 1

1612, 1

226, 1

1332, 1

1333, 1

1334, 1

1335, 1

1062, 1

1063, 1

4831, 1

3037, 1

1070, 1

1071, 1

3426, 1

3427, 1

3938, 1

1986, 1

5178, 1

3725, 1

5344, 1

5345, 1

3918, 1

5555, 1

2071, 1

2072, 1

1452, 1

1453, 1

1454, 1

1455, 1

573, 1

3100, 1

4720, 1

2458, 1

2464, 1

2669, 1

2670, 1

2671, 1

5835, 1

2611, 1

1157, 1

1160, 1

1163, 1

1165, 1

1168, 1

1171, 1

1172, 1

307, 1

309, 1

311, 1

312, 1

313, 1

314, 1

316, 1

317, 1

319, 1

320, 1

153, 1

3168, 1

224, 1

3619, 1

278, 1

280, 1

282, 1

283, 1

284, 1

285, 1

286, 1

287, 1

288, 1