**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SANTA CATARINA – CÂMPUS FLORIANÓPOLIS**
**DEPARTAMENTO ACADÊMICO DE ELETROTÉCNICA**
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**
**MESTRADO PROFISSIONAL EM SISTEMAS DE ENERGIA ELÉTRICA**

**ADRISSON CONSONI FLORIANO**

# PROPOSIÇÃO DE METODOLOGIA PARA TRANSCRIÇÕES AUTOMATIZADAS DE ÁUDIOS PARA CENTROS DE OPERAÇÃO DO SETOR ELÉTRICO

**FLORIANÓPOLIS, 2021.**

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SANTA CATARINA – CÂMPUS FLORIANÓPOLIS**
**DEPARTAMENTO ACADÊMICO DE ELETROTÉCNICA**
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**
**MESTRADO PROFISSIONAL EM SISTEMAS DE ENERGIA ELÉTRICA**

**ADRISSON CONSONI FLORIANO**

# PROPOSIÇÃO DE METODOLOGIA PARA TRANSCRIÇÕES AUTOMATIZADAS DE ÁUDIOS PARA CENTROS DE OPERAÇÃO DO SETOR ELÉTRICO

Dissertação de mestrado submetida ao Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica.

Professor Orientador: Sérgio Luciano Ávila, D. Eng.

Coorientador: Rubipiara Cavalcante Fernandes, D. Eng.

**FLORIANÓPOLIS, 2021.**

Ficha de identificação da obra elaborada pelo autor.

CONSONI FLORIANO, ADRISSON
PROPOSIÇÃO DE METODOLOGIA PARA TRANSCRIÇÕES AUTOMATIZADAS DE ÁUDIOS PARA CENTROS DE OPERAÇÃO DO SETOR ELÉTRICO / ADRISSON CONSONI FLORIANO; orientação de Sérgio Luciano Ávila; coorientação de Rubipiara Cavalcante Fernandes. - Florianópolis, SC, 2021.
95 p.
Dissertação (Mestrado) - Instituto Federal de Santa Catarina, Câmpus Florianópolis. Mestrado Profissional em Sistemas de Energia Elétrica. Departamento Acadêmico de Eletrotécnica.
Inclui Referências.

1. Transcrição. 2. Vocabulário Personalizado. 3. Setor Elétrico. 4. Centro de Controle. 5. Audio. I. Luciano Ávila, Sérgio. II. Cavalcante Fernandes, Rubipiara . III. Instituto Federal de Santa Catarina. IV. PROPOSIÇÃO DE METODOLOGIA PARA TRANSCRIÇÕES AUTOMATIZADAS DE ÁUDIOS PARA CENTROS DE OPERAÇÃO DO SETOR ELÉTRICO.

# PROPOSTA DE METODOLOGIA PARA TRANSCRIÇÕES AUTOMATIZADAS DE ÁUDIOS PARA CENTROS DE OPERAÇÃO DO SETOR ELÉTRICO

**ADRISSON CONSONI FLORIANO**

Este trabalho foi julgado adequado para obtenção do título de Mestre em Engenharia Elétrica e aprovado na sua forma final pela banca examinadora do Mestrado Profissional em Sistemas de Energia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina.

Florianópolis, 30 de junho de 2021.

Banca Examinadora:

Documento assinado digitalmente
SERGIO LUCIANO AVILA
Data: 19/08/2021 15:24:15-0300
CPF: 020.743.929-07

Prof. Sérgio Luciano Ávila,
D. Eng – IFSC

Assinado de forma digital por Rubipiara Fernandes
Dados: 2021.08.24 11:22:54 -03'00'

Prof. Rubipiara Cavalcante Fernandes,
D. Eng – IFSC

Luciana Moura Nabarrete
Analista de Sistemas – ENGIE Brasil

Maurício Barbosa de Camargo Salles

Assinado de forma digital por Maurício Barbosa de Camargo Salles
DN: cn=Maurício Barbosa de Camargo Salles, o=Universidade de São Paulo, ou=Escola Politécnica, email=mausalles@usp.br, c=BR
Dados: 2021.08.24 10:21:31 -03'00'

Prof. Maurício Barbosa de Camargo Salles,
D. Eng – USP

Assinado de forma digital por EDISON ANTONIO CARDOSO ARANHA NETO:00583546978
Dados: 2021.08.24 12:59:27 -03'00'

Prof. Edison Antonio Cardoso Aranha Neto,
D. Eng – IFSC

# AGRADECIMENTOS

Agradeço ao IFSC como instituição e ao corpo docente do programa de Mestrado em Sistemas de Energia, por disponibilizarem o programa, possibilitando aos profissionais que atuem no setor e que sempre tiveram interesse pela área acadêmica, a possibilidade de conciliar as atividades. A aproximação entre empresas e academia é benéfica para ambas, para os profissionais e para a sociedade.

Gostaria de reconhecer o esforço e contribuição dos professores envolvidos neste trabalho. Ao professor e orientador Sérgio Luciano Ávila, pela presteza, disponibilidade e pelo entusiasmo com o tema. Sem ele, esta dissertação não poderia ter sido defendida em tempo, nem teria a qualidade apresentada. Ao professor e coorientador Rubipiara Cavalcante Fernandes, mentor desde o ano de 2006, quando cursei as primeiras disciplinas do então Curso Tecnólogo em Gestão de Sistemas de Energia, e que acreditou que o tema poderia ser objeto de estudo da dissertação. Muito obrigado!

Registro minha gratidão à ENGIE pela cessão dos dados e por ser uma empresa que incentiva a inovação e iniciativa dos colaboradores, propiciando um ambiente sem o qual o tema desta dissertação e a motivação para cursar este mestrado, não existiriam. Expresso aqui também minha gratidão aos colegas de empresa que, conscientes ou não, ajudaram bastante durante estes dois anos de mestrado. Seja contribuindo diretamente para o entendimento de alguns temas, segurando as pontas nos momentos de maior demanda do curso, ou apenas partilhando uma conversa bem humorada nos momentos mais difíceis, foram fundamentais para esta conquista.

Não poderia deixar de agradecer à minha família, fonte de apoio incondicional e motivação de todas as minhas conquistas. Iraci, Gilmar, Gabriela e Guilherme, muito obrigado! Vocês são parte disso.

Por fim, gostaria de agradecer àquela que esteve comigo durante esta jornada. Muito obrigado pelas palavras de conforto nas derrotas, motivação nas horas difíceis e parceria nas comemorações. À minha companheira Stephanie, o mais especial dos agradecimentos.

# RESUMO

O setor elétrico passa por importante evolução quanto à adoção de novas tecnologias que, embora desejáveis e inevitáveis, acrescentam complexidade operacional ao sistema. Graças à essência de suas atividades e à visão sistêmica necessária para sua execução, os centros de operação são concentradores naturais de informação. Sejam valores operacionais de sensores, ou diretivas oriundas de órgãos reguladores, os dados se cruzam nos centros de operação, onde são armazenados em distintos formatos e utilizados para diferentes aplicações. No tocante às chamadas por áudio, grande parte da comunicação entre as equipes de tempo real dos agentes com operadores do ONS é feita via canal de voz. Essas gravações são armazenadas, gerando grande volume de dados e consultadas sob demanda. Elas servem como evidências para confirmação de eventos, procedimentos, manutenções, alteração dos estados operativos, ou quaisquer intervenções nas instalações elétricas. Na prática, esses arquivos de áudio contêm muitas outras informações relevantes para a operação do sistema e faturamento dos agentes, mas acabam sendo subutilizados devido à dificuldade de extração, rastreabilidade e associação destes áudios com outros registros dos demais sistemas operativos. Em virtude da evolução e da popularização de técnicas de manipulação e tratamento de grandes quantidades de dados, tornou-se possível extrair informações estruturadas dos arquivos de áudio das empresas de geração, transmissão e distribuição. A principal contribuição deste trabalho é proposição de uma metodologia para o desenvolvimento de um vocabulário estruturado próprio ao setor elétrico. Para tal, utiliza-se técnicas de *Speech-to-Text* para fazer a transcrição dos registros de comunicações das salas de controle, com processamento de linguagem natural, que leva em consideração tanto o vocabulário técnico, quanto a pluralidade de sotaques e regionalismos do território brasileiro. Ainda, discute-se métricas para a avaliação da qualidade do vocabulário obtido. A metodologia aqui desenvolvida torna a busca por evidências na operação mais dinâmica e eficiente. Por fim, outra contribuição deste trabalho é criar os alicerces para aplicações mais sofisticadas, que façam uso dos dados estruturados, oriundos dos áudios analisados.

**Palavras-chave**: Transcrição. Áudio. Vocabulário Personalizado. Setor Elétrico. Centro de Controle.

# ABSTRACT

The power sector is undergoing an evolution regarding the new technologies adoption that, although required and unavoidable, add operational complexity to the system. Thanks to the essence of their activities and the systemic vision necessary for their execution, the operation centers are natural information hubs. Whether operating values from sensors or directives from regulatory bodies, the data is crossed in the operation centers, stored in different formats, and used for many applications. Concerning audio calls, a large part of the communication between the agent's real-time teams and ONS operators is done by voice channel. These recordings are stored, generating a large volume of data, and consulted on demand. They serve as evidence for confirming events, procedures, maintenance, changes in operating states, or any interventions in electrical installations. In practice, these audio files contain much other information relevant to the system's operation and agents' billing. However, it ends up being unused due to the difficulty of extracting, tracking, and associating these audios with other records from other operating systems. Due to the evolution and popularization of techniques for handling and processing large amounts of data, it has become possible to extract structured information from the audio files of generation, transmission, and distribution companies. The main contribution of this work is a methodology for a structured vocabulary specific to the power sector. Speech-to-Text techniques are used to transcribe the communication records of the control rooms, with natural language processing, which takes both the technical vocabulary and the plurality of accents and regionalisms of the Brazilian language. Furthermore, metrics for assessing the vocabulary quality obtained are discussed. The methodology developed seeks evidence of a more dynamic and efficient operation.



**Keywords**: Speech-to-Text. Audio. Custom Vocabulary. Power Sector. Control Room.

# LISTA DE FIGURAS

Figura 1 - Esquemático do escopo da dissertação....11
Figura 2 - Hierarquia do sistema de comunicação do ONS e agentes....18
Figura 3 - Comparação de WER e KER....24
Figura 4 - Exemplo de NER - transcrição de referência....29
Figura 5 - Exemplo de NER - transcrição de hipótese....29
Figura 6 - Média e desvio padrão do WER pelo número de transcrições....36
Figura 7 - Histograma das ligações por ramais de origem....37
Figura 8 - Histograma das ligações por ramais de destino....38
Figura 9 - Distribuição das ligações ao longo do dia....39
Figura 10 - Histograma de duração das ligações....40
Figura 11 - Comparação entre soluções de transcrição....49
Figura 12 - Transcrição utilizando vocabulários personalizados parciais....54
Figura 13 - Arquitetura da metodologia proposta....63
Figura 14 - Metodologia de tratamento das transcrições - WER, KER e NER....65

# LISTA DE TABELAS

Tabela 1 - Exemplo de cálculo do WER ..... 20
Tabela 2 - Lista de Palavras-Chave ..... 26
Tabela 3 - Exemplo de *over stemming* ..... 28
Tabela 4 - Exemplo de *under stemming* ..... 28
Tabela 5 - Exemplo de reconhecimento de entidades utilizando NER ..... 30
Tabela 6 - Comparação WER entre as soluções – Arquivos WAV ..... 44
Tabela 7 – Comparação WER entre as soluções – Arquivos MP3 ..... 44
Tabela 8 – Comparação KER entre as soluções ..... 45
Tabela 9 – Comparação WER com remoção de *stop words* ..... 46
Tabela 10 – Comparação KER com s*temming* ..... 46
Tabela 11 – Comparação NER entre localidades ..... 48
Tabela 12 – Comparação NER entre organizações ..... 48
Tabela 13 – Comparação NER entre todas as entidades ..... 48
Tabela 14 - Modelando diferentes pronúncias no vocabulário personalizado ..... 53
Tabela 15 - Termos comuns aos vocabulários personalizados ..... 55
Tabela 16 - Vocabulário personalizado do COG ..... 56
Tabela 17 - Vocabulário personalizado de usinas hidrelétricas ..... 57
Tabela 18 - Vocabulário personalizado de usinas termelétricas ..... 58
Tabela 19 - Vocabulário personalizado de usinas fotovoltaicas ..... 58
Tabela 20 - Vocabulário personalizado das chamas em geral ..... 59

# LISTA DE ABREVIATURAS E SIGLAS

AFI - Alfabeto Fonético Internacional

ANEEL - Agência Nacional de Energia Elétrica

ASR - *Automatic Speech Recognition* (Reconhecimento Automático de Fala)

AWS - *Amazon Web Services*

BLEU - *Bilingual evaluation understudy*

CCEE - Câmara de Comercialização de Energia Elétrica

CDER - *Cover Disjoint Error Rate*

COG - Centro de Operação da Geração

COS - Centro de Operação de Sistemas

CPFL - Companhia Paulista de Força e Luz

DAU - Desligado automaticamente por atuação de sistema de proteção

DCO - Desligado por conveniência operativa

DESSEM - Modelo de Despacho Hidrotérmico de Curto Prazo

GCP - *Google Cloud Platform*

GT - *Ground Truth* (Verdade Fundamental)

IFSC - Instituto Federal de Santa Catarina

KER - *Keyword Error Rate*

LCC - Ligado como compensador síncrono, por solicitação do agente

LCS - Ligado como compensador síncrono, por solicitação do sistema

LIG - Ligado como gerador

LOC - Localização

MACS - *Microsoft Azure Cognitive Services*

MER - *Match Error Rate*

METEOR - *Metric for Evaluation of Translation with Explicit Ordering*

MPO - Manual de Procedimento da Operação

NER - *Named Entity Recognition*

NIST - *National Institute of Standards and Technology*

NLP - *Natural Language Processing* (Processamento de Linguagem Natural)

NLTK - *Natural Language Toolkit*

ONS - Operador Nacional do Sistema Elétrico

ORG - Organização

PER - *Position Independent Word Error Rate*

POS - *Part of Speech*

PR - Procedimentos de Rede

RAO - Relatório de Análise de Operação

RAP - Relatório de Análise de Perturbação

RIL - *Relative Information Lost*

RIP - *Relative Information Preserved*

RO - Relatório de Análise de Ocorrência

SCADA - *Supervisory control and data acquisition* (Sistema Supervisório)

SIN - Sistema Interligado Nacional

StT - *Speech-to-Text* (Transcrição de Áudio Para Texto).

TER - *Translation Error Rate*

UFSC - Universidade Federal de Santa Catarina

URGS - Universidade Federal do Rio Grande do Sul

USP - Universidade de São Paulo

WER - *Word Error Rate*

WIL - *Word Information Lost*

WIP - *Word Information Preserved*

# SUMÁRIO


**1 INTRODUÇÃO ..... 9**
**1.1 Objetivos ..... 12**
**1.1.1 Objetivos Específicos ..... 13**
**1.2 Sobre o autor ..... 13**
**2 BASE CONCEITUAL PARA TRANSCRIÇÃO DE ÁUDIOS NO SETOR ELÉTRICO ..... 15**
**2.1 Operação de ativos no Setor Elétrico ..... 16**
**2.2 Requisitos operacionais normativos ..... 18**
**2.3 Medição da qualidade das transcrições ..... 19**
2.3.1 *Word Error Rate* (WER) ..... 19
2.3.2 WER por componente ..... 22
2.3.3 *Keyword Error Rate* (KER) ..... 23
2.3.4 *Named Entity Recognition* (NER) ..... 28
2.3.5 Outras métricas e estratégias ..... 31
**2.4 Tratamento de arquivos de áudio ..... 33**
***2.5* Definição do *ground truth* ..... 35**
2.5.1 Análise dos metadados ..... 36
**2.6 Comparação entre ferramentas de transcrição ..... 40**
2.6.1 Critérios de Comparação ..... 43
**2.7 Criação dos vocabulários personalizados ..... 51**
2.7.1 Vocabulários Parciais ..... 53
**3 PROPOSIÇÃO DE METODOLOGIA ..... 61**
**3.1 Arquitetura proposta ..... 61**
**3.2 Avaliação da qualidade das transcrições ..... 64**
**3.3 Potenciais aplicações e futuros trabalhos ..... 66**
3.3.1 Busca por palavras-chave ..... 66
3.3.2 Classificação de chamadas ..... 66
3.3.3 Avaliação da qualidade da comunicação ..... 67
3.3.4 Passagens de turno ..... 67
3.3.5 Mudança automática de estados operacionais ..... 68
3.3.6 Indicadores ..... 69
**4 CONCLUSÃO ..... 70**
**REFERÊNCIAS ..... 72**
**APÊNDICES ..... 78**
APÊNDICE A – AUTORIZAÇÃO DE DIVULGAÇÃO DE INFORMAÇÕES ..... 79
APÊNDICE B – CÁLCULO DAS MÉTRICAS WER, MER, WIL, WIP ..... 80
APÊNDICE C – PRÉ-PROCESSAMENTO DAS TRANSCRIÇÕES ..... 82
APÊNDICE D – CÁLCULO DA MÉTRICA KER ..... 84
APÊNDICE E – REMOÇÃO DE *STOP WORDS* ..... 86
APÊNDICE F – APLICAÇÃO DO *STEMMING* ..... 87
APÊNDICE G – CÁLCULO DAS MÉTRICAS NER ..... 88
APÊNDICE H – PROCESSAMENTO DAS TRANSCRIÇÕES ..... 91

# 1 INTRODUÇÃO

O setor elétrico brasileiro vem sendo reestruturado a fim de atender às necessidades de uma sociedade em constante desenvolvimento e modernização. Os profissionais que atuam no setor buscam soluções para problemas em diferentes frentes:

a) sob a ótica da operação de ativos do setor elétrico, a teleassistência de usinas e subestações tem se tornado solução recorrente entre os principais agentes do setor. Eles devem atender aos requisitos dispostos nos Procedimentos de Rede (PR), para contemplar as necessidades desta nova modalidade de operação (ONS, 2020);
b) no que se refere ao planejamento energético, as fontes intermitentes estão aumentando a sua relevância na matriz elétrica nacional e apesar das vantagens econômicas e ambientais, suas características não despacháveis demandam uma operação cada vez mais eficiente (PINHEIRO, 2020; BJARGHOV, *et al.*, 2021; CUISINIER, *et al.*, 2020);
c) do ponto de vista econômico, em busca de eficiência e competitividade, o preço semanal da energia e a programação do despacho estão em processo de migração para intervalos semi-horários (FERNANDEZ, 2019). Desde 2020, o Operador Nacional do Sistema Elétrico (ONS) vem utilizando o Modelo de Despacho Hidrotérmico de Curto Prazo (DESSEM) (ONS, 2021) na programação da geração diária, com patamares de 30 minutos. Apesar de benefícios como transparência e reprodutibilidade dos processos e redução da ocorrência de geração fora da ordem de mérito, estas mudanças afetam as equipes de operação de tempo real, que precisam ajustar a geração com maior frequência; e as equipes de pré-operação, que necessitam acordar o programa de geração, junto ao ONS, para o dia seguinte (ONS, 2021);
d) a geração distribuída está cada vez mais presente e os consumidores-produtores são mais numerosos a cada dia (ANEEL, 2020). Além das vantagens como geração de energia próxima à carga e diminuição de perdas técnicas, atendimento às cargas isoladas e utilização de fontes renováveis, estas novas tecnologias demandam novas soluções de controle dos níveis de

tensão, proteção de redes de distribuição e previsão de carga (DIAS; BOROTNI; HADDAD, 2015; YANG, *et al.*, 2020; CHEN, *et al.*, 2021).

Todas estas iniciativas, embora desejáveis e inevitáveis, acrescentam complexidade operacional ao sistema. Além de manter os procedimentos e ferramentas já consolidados, o desafio da atividade de operação de sistemas elétricos é não se tornar ela própria um fator limitante para a incorporação dessas evoluções. Para tal, ela precisa se reinventar, de forma a assumir o seu protagonismo como viabilizadora do setor e responder às mudanças com primor técnico.

Graças à essência de suas atividades e à visão sistêmica necessária para sua execução, os centros de operação são concentradores naturais de informação. Sejam valores operacionais de sensores em campo, ou informações gerenciais provenientes de outros sistemas ou órgãos reguladores, os dados se cruzam nos centros de operação, onde são armazenados em distintos formatos e utilizados para diferentes aplicações.

No tocante às chamadas por áudio, grande parte da comunicação entre as equipes de tempo real dos agentes com os técnicos em campo ou operadores do ONS é feita via canal de voz. Essas gravações são armazenadas (ONS, 2020), gerando grande volume de dados e consultadas sob demanda, servindo como evidências para a elaboração dos Relatórios de Análise de Ocorrência (RO), dos Relatórios de Análise da Operação (RAO) (ONS, 2020) e dos Relatórios de Análise de Perturbações (RAP) (ONS, 2020). Na prática, esses arquivos de áudio contêm muitas outras informações relevantes para a operação do sistema e faturamento dos agentes, mas acabam sendo subutilizados devido à dificuldade de extração (BATES; WEISCHEDEL, 1993), rastreabilidade e associação destes áudios com outros registros dos demais sistemas operativos.

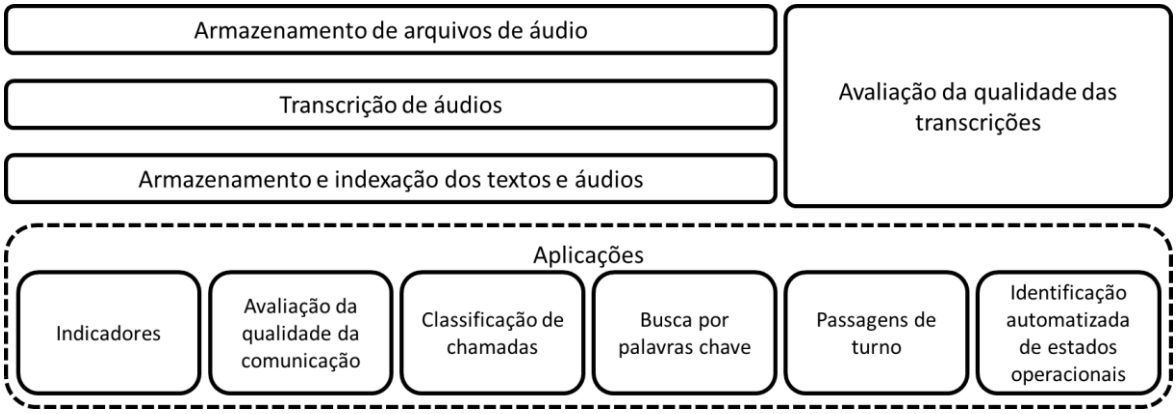
Nos últimos anos, entretanto, a evolução e a popularização de técnicas de manipulação e de tratamento de grandes quantidades de dados tornou possível extrair informações estruturadas dos arquivos de áudio (AUDHKHASI, *et al.*, 2017). Várias metodologias vêm sendo desenvolvidas para transformar a língua falada em texto, podendo ser agrupadas e chamadas de técnicas de *Speech-to-Text* (StT) (KANO; SAKTI; NAKAMURA, 2020; SCHARENBORG, *et al.*, 2020).

Este trabalho propõe usar técnicas de StT para realizar a transcrição dos registros de comunicações das salas de controle, com processamento de linguagem natural (NLP), que leve em consideração tanto o vocabulário técnico do setor elétrico, quanto a pluralidade de sotaques e regionalismos do território nacional.

Esta dissertação é escrita sob a ótica do segmento de geração, devido à expertise do autor e à sua proximidade ao objeto de estudo. Para desenvolvimento deste trabalho, foram utilizadas amostras de áudio e documentações da ENGIE e do NOS. Os dados utilizados são de propriedade da ENGIE e a sua utilização foi autorizada exclusivamente para fins acadêmicos. O termo de autorização de uso das informações da empresa pode ser encontrado no APÊNDICE A.

O trabalho utiliza como insumo os dados de áudio armazenados do Centro de Operação de Sistemas (COS) da ENGIE. Sobre essa massa de dados, é primeiro realizada uma análise das melhores práticas para transcrição; em seguida são feitas comparações entre soluções de mercado de StT; e por fim tem-se a criação de vocabulário personalizado para melhoria da qualidade das transcrições (Capítulo 2). Trata-se de um trabalho de base, cujos resultados (transcrições) são armazenados em formato de texto e indexadas com os áudios originais, podendo servir de insumo para diversas outras aplicações, como classificação de chamadas, criação de indicadores operacionais, identificação de mudança de estados operacionais, entre outras possibilidades (Capítulo 3). A Figura 1 exibe o escopo desta dissertação.

**Figura 1 - Esquemático do escopo da dissertação**




Fonte: Elaboração própria (2021).

Embora sejam desafios técnicos estimulantes, as aplicações assumem os dados transcritos e prontos para uso, sem aprofundar em como foram obtidos. Elas

não são escopo deste trabalho, sendo discutidas para fins de contextualização e inspiração para trabalhos vindouros.

Em problemas reais, diferentes aplicações podem ter diferentes sensibilidades quanto à qualidade das transcrições. Soluções de classificação de chamadas, por exemplo, normalmente utilizam técnicas que simplificam e agrupam palavras, entre outras formas de degradação da mensagem original. Ou seja, para garantir uma melhor taxa de classificação, é interessante que a mensagem analisada seja concisa: “aumentar”, “volts” e “Itá” são termos suficientes para classificar, com um bom grau de certeza, uma chamada, mesmo que as mensagens sejam ilegíveis para leitores humanos. Para estes casos, a exatidão da transcrição pode e deve ser sacrificada, para otimizar a classificação.

Outras aplicações, como a avaliação da qualidade da comunicação por exemplo, dependem fortemente da qualidade das transcrições. Para este tipo de aplicação, a identificação de uma repetição durante as transcrições pode significar tanto uma falha de comunicação, quanto o atendimento satisfatório a um critério obrigatório. O fato de um mesmo interlocutor repetir, consecutivamente, a mesma mensagem, pode ser indício de falta de clareza na comunicação. Em contrapartida, o ato de ser repetida por interlocutores diferentes, é indicativo de que os operadores confirmaram, satisfatoriamente, o envio e recebimento da mensagem.

Nesta dissertação serão apresentadas brevemente algumas aplicações que podem utilizar os dados transcritos, mas o aprofundamento na implementação e funcionamento de cada uma, não será explorado. Este trabalho, entretanto, tem por objetivo transcrever, com a maior fidelidade possível, os áudios originais, delegando às aplicações a decisão de usar ou não todas as informações entregues.

## 1.1 Objetivos

O objetivo desta dissertação é propor uma metodologia de transcrição automatizada de áudio que faça uso de um vocabulário personalizado, adaptado ao setor elétrico e que considere os regionalismos do território nacional.

### 1.1.1 Objetivos Específicos

Além de desenvolver uma proposição de metodologia para transcrição de áudios do setor elétrico, este trabalho tem outros objetivos, que são utilizados para embasar o objetivo principal e podem servir como referência para trabalhos relacionados.

a) Converter uma exigência regulatória de armazenamento de arquivos de áudio de comunicação, reaproveitando-os como fonte de dados para a operação, aplicando técnicas de aprendizado de máquina de forma a transformar os arquivos brutos em inteligência operacional, através da transcrição dos áudios e estruturação das informações;

b) Definir métricas de quantificação da qualidade das transcrições realizadas, que permita avaliar as evoluções e a comparação entre diferentes vocabulários, técnicas e ferramentas;

c) Realizar a curadoria das informações relevantes para a utilização de arquivos de áudios em centros de controle do setor elétrico, sob as óticas regulatória e técnica.

## 1.2 Sobre o autor

Engenheiro eletricista graduado pela Universidade Federal de Santa Catarina (UFSC) e discente do Mestrado Profissional em Sistemas de Energia no Instituto Federal de Santa Catarina (IFSC). Atua desde 2009 com soluções tecnológicas para o setor elétrico. Aficionado por tecnologia, tem especial interesse por soluções inovadoras e criativas. Trabalha desde 2018 na gerência de operações da ENGIE, em Florianópolis, de onde são controladas remotamente 47 usinas do grupo, em todas as regionais do país, totalizando aproximadamente 4,5 GW de capacidade instalada.

Ao entrar em contato com as equipes de operação de tempo real do COS da ENGIE, percebeu algumas das suas dificuldades e, aproveitando a estratégia de digitalização da empresa, idealizou um Hackathon. O evento teve duração de um fim

de semana, no qual foram propostos desafios técnicos e os competidores desenvolveram protótipos de soluções. O evento foi realizado no Rio de Janeiro, em julho de 2019, na sede da Radix Engenharia e Software e contou com desafios de ENGIE e da Companhia Paulista de Força e Luz (CPFL), que apresentaram problemas reais das companhias, nos segmentos de Geração e Distribuição de energia, respectivamente. Um dos desafios apresentados pela equipe da ENGIE foi a dificuldade de rastreabilidade da comunicação, via áudio, entre os centros de operação do agente e do ONS. As soluções desenvolvidas durante o evento foram tão promissoras, com aplicações tanto no segmento de geração quanto no de distribuição, que ENGIE e CPFL decidiram investir na transcrição de áudios para centros de controle do setor elétrico e iniciou-se um projeto cooperado, com financiamento do Programa de Pesquisa e Desenvolvimento da Agência Nacional de Energia Elétrica (ANEEL), sob código PD-00063-3071/2020.

A presente dissertação é peça a parte deste projeto ANEEL. Embora sejam trabalhos complementares, inseridos no mesmo contexto, os desenvolvimentos desta dissertação foram elaborados pelo autor durante a execução do projeto e exclusivamente para esta dissertação. Alguns resultados já foram incorporados ao projeto ANEEL.

Ao longo do desenvolvimento desta dissertação foram submetidos os seguintes artigos:

I. Transcrição de áudios em centros de operação e correlação com dados da rede: transformando gravações brutas em inteligência operacional; XVI EDAO; junho de 2021. **Apresentado**.
II. *Multi-label classification of voice calls from power plant operation centers; IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*; setembro 2021. **Classificado**.
III. Transcrição automática de áudios em Centros de Operação; 7º Seminário de Operação ENGIE; outubro de 2021. **Classificado**.
IV. Transcrição de áudios para o setor elétrico: criação de vocabulário especializado para utilização em soluções de mercado; X SENOP; novembro de 2021. **Classificado**.

# 2 BASE CONCEITUAL PARA TRANSCRIÇÃO DE ÁUDIOS NO SETOR ELÉTRICO

Neste capítulo são exploradas as bases conceituais para proposição de uma metodologia para transcrições de áudios dos centros de operação do setor elétrico. No primeiro momento é descrito o fluxo de informações de áudio entre centros de controle do ONS e dos agentes, considerando operações de ativos remotamente ou localmente. Em seguida são apresentados os requisitos operacionais referentes à gravação de áudio.

Logo após é realizado um aprofundamento nas métricas de quantificação da qualidade das transcrições. Esta etapa define os conceitos e os métodos de avaliação utilizados na subsequente comparação de diferentes ferramentas e vocabulários de transcrição. Além de um apanhado na bibliografia mais pertinente, são apresentadas as métricas utilizadas para a avaliação das transcrições de áudios dos centros de controle do setor elétrico, em português, e contemplam algumas das contribuições mais importantes deste trabalho.

Na sequência são exploradas algumas técnicas de tratamento da qualidade dos arquivos de áudio. O objetivo é aumentar a qualidade das transcrições, aplicando filtros e tratamentos para remoção de ruídos e outros fatores que degradam o desempenho dos modelos de transcrição.

Dando continuidade ao capítulo, é realizada uma comparação entre cinco ferramentas de transcrição consolidadas no mercado: *Amazon Web Services* (AWS); Dígitro; *Google Cloud Platform* (GCP); IBM Watson e *Microsoft Azure Cognitive Services* (MACS). Nesta seção são aplicadas diferentes métricas a fim de elencar as soluções com melhor desempenho de transcrições no contexto deste trabalho. Além disso, será explorada a capacidade das soluções AWS e GCP de utilizarem vocabulários personalizados, mesmo em nível superficial.

Por fim, são investigadas com profundidade as possibilidades de personalizações de vocabulários. É apresentada uma visão geral do funcionamento e da configuração dos arquivos de vocabulários fonéticos da solução escolhida, AWS, bem como as boas práticas. Além disso, serão apresentadas as estratégias utilizadas para melhorar a qualidade das transcrições, considerando nomes próprios de ativos, termos técnicos e outros jargões.

Para as ações exploratórias e desenvolvimentos, foi utilizado como *ground truth* (GT) um conjunto de 200 áudios e suas respectivas transcrições manuais. G*round truth* (verdade fundamental) é um termo usado para se referir a informações que são conhecidas como reais ou verdadeiras, fornecidas por observação direta e medição, em oposição às informações fornecidas por inferência. No contexto desta dissertação, o GT é formado pelas transcrições manuais dos áudios reais, realizadas pelo autor.

Os arquivos são dados reais, gravados no COS da ENGIE. Foram escolhidas amostras em formato WAV com *bit rate* de 256 kbps e MP3 com *bit rate* de 64 kbps. *Bit rate* significa fluxo de transferência de *bits* e está vinculado à qualidade dos arquivos de áudio (TORCOLI; KASTNER; HERRE, 2021). Além dos áudios em si, o gravador utilizado pela ENGIE disponibiliza alguns metadados que serão explorados neste trabalho, como data e instante da chamada, duração em segundos, ramal de origem e ramal de destino.

Os algoritmos utilizados foram desenvolvidos em *Python*. Trata-se de uma linguagem de programação interpretada que dispões de milhares de bibliotecas especializadas nas mais diferentes aplicações (PYTHON SOFTWARE FOUNDATION, 2021). Neste trabalho foram utilizadas bibliotecas do *Python* que calculam os indicadores de transcrição mais difundidos, permitem a manipulação de textos e contam com funções de tratamento de áudios. Estas vantagens economizam esforços no desenvolvimento de funcionalidades, permitindo que o autor possa focar nos aspectos referentes ao setor elétrico.

## 2.1 Operação de ativos no Setor Elétrico

No Brasil, as atividades de coordenação e controle da operação da geração e da transmissão de energia elétrica do Sistema Interligado Nacional (SIN) são realizadas pelo ONS. Fazem parte do SIN todas as instalações eletricamente interligadas, responsáveis pelo suprimento de energia elétrica. Consideradas suas atribuições legais, todas as atividades do ONS são baseadas em regras, critérios e procedimentos técnicos, e estão organizados nos módulos dos Procedimentos de Rede (ONS, 2008). O atendimento regulatório em conformidade com os

procedimentos de rede é condição indispensável para qualquer tecnologia que almeje compor o portifólio de soluções do setor elétrico.

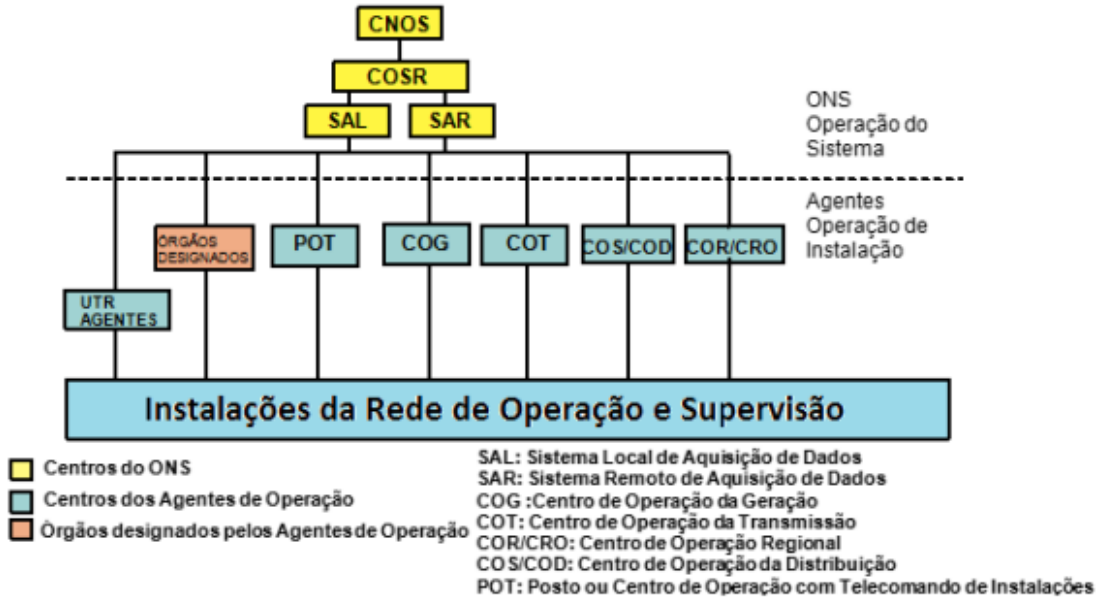
Em conformidade com a Lei 10.848/04 e com o Decreto 5.081/04, desde 2004 o ONS é responsável, dentre outras atribuições, pelo planejamento e programação da operação, bem como o despacho centralizado da geração, além da supervisão e coordenação dos centros de operação de sistemas elétricos (ONS, 2008). Todas as usinas em território nacional são classificadas de acordo com a sua modalidade operacional, em Tipo I, II ou III (ONS, 2021)

São classificadas como tipo III as usinas conectadas fora da Rede Básica e que não causam impactos na operação eletroenergética do SIN ou empreendimentos de autoprodução conectados na Rede Básica, cuja demanda seja permanentemente maior que a geração. As usinas classificadas na modalidade de operação Tipo III não possuem programação nem despacho centralizados e, por este motivo, não possuem relacionamento operacional com o ONS (ONS, 2021), logo, não fazem parte do escopo deste trabalho.

As usinas dos tipos I e II, que possuem relacionamento direto com o ONS, devem dispor de serviços de comunicação de voz que suportem as atividades de normatização, pré-operação, operação em tempo real, pós-operação, apoio e coordenação de telecomunicações, planejamento e programação da operação (ONS, 2020).

O ONS se comunica por voz com o centro de operação dos agentes ou, quando não existem, diretamente com suas instalações (ONS, 2020). Os centros de operação do ONS e dos agentes devem dispor de recursos de gravação de voz e os arquivos dos registros devem ser armazenados por um período mínimo de dois anos (ONS, 2020). A Figura 2 exibe configurações possíveis para os serviços de telecomunicações de voz, considerando os centros de operação do ONS e agentes.

**Figura 2 - Hierarquia do sistema de comunicação do ONS e agentes**




Fonte: ONS (2020).

## 2.2 Requisitos operacionais normativos

Como todos os aspectos técnicos relacionados à operação de sistemas elétricos, os requisitos de chamadas por voz e gravações são definidos pelos PR. No que tange à gravação das chamadas, os procedimentos de rede exigem que os centros de operação do ONS e dos agentes disponham de recursos de gravação de voz e, também, para os centros de operação dos agentes, sistema de gravação das comunicações operacionais com as suas instalações (ONS, 2020).

Sobre o prazo de armazenamento dos arquivos, os PR especificam dois meses para gravações convencionais e dois anos para gravações que foram objeto de relatórios e análises (ONS, 2020). Na prática, os custos de armazenamento digital são cada vez mais acessíveis (JADON; BHADORIA; TOMAR, 2015). Assim, considerando as complexidades operacionais de selecionar os arquivos que devem ser mantidos e a decisão de dispor do histórico e rastreabilidade das ocorrências, a ENGIE opta por armazenar todas as ligações, por pelo menos quatro anos. Além disso, as informações gravadas devem ser prontamente disponibilizadas para a ANEEL e para o ONS, quando solicitadas. No caso de perda de gravação ou falha no gravador do agente, caso seja necessário para auxiliar a análise de ocorrências no

SIN, o ONS poderá disponibilizar ao agente as gravações entre os seus centros e do agente (ONS, 2020).

Foi identificado que a ENGIE possui um banco de dados de áudios com registros desde 2017, atualmente com aproximadamente 90 GB, distribuídos em mais de 550 mil arquivos MP3 e com duração acumulada de 364 dias. Estes números são incrementados a uma taxa média de 700 arquivos, totalizando 9 horas de gravações e 100 MB de armazenamento por dia. Devido à dificuldade de processar arquivos de áudios e transformar informações de linguagem natural em dados estruturados (BATES; WEISCHEDEL, 1993), essa massa de dados é armazenada e consultada sob demanda, durante as elaborações dos Relatórios de Análise RO, RAO (ONS, 2020) e RAP (ONS, 2020).

## 2.3 Medição da qualidade das transcrições

Um dos objetivos deste trabalho é desenvolver um dicionário fonético adaptado ao setor elétrico, de forma que as transcrições sejam tão fidedignas quanto possível aos áudios originais. Nesta seção são exploradas diferentes abordagens com o propósito de melhorar a qualidade das transcrições. Para que esta estratégia seja exitosa, entretanto, é necessário definir um critério objetivo para quantizar a qualidade das transcrições. Definir uma métrica adaptada ao contexto é imprescindível para comparar diferentes soluções e avaliar evoluções de modelos. Trata-se de um dos principais desafios deste trabalho.

Nesta seção são apresentadas as métricas mais difundidas na literatura e adaptações desenvolvidas para o contexto desta dissertação. Para fins de organização, todas as métricas de quantização da qualidade das transcrições são apresentadas antes das soluções. Esta estratégia permite avaliar o impacto de cada estratégia adotada na qualidade das transcrições.

### 2.3.1 *Word Error Rate* (WER)

O WER é uma métrica consolidada nos contextos de transcrição de linguagem falada (WANG; ACERO; CHELBA, 2003) e tradução automática de textos (POPOVIC; NEY, 2007). O WER mede a diferença entre duas sequências de

palavras. Para quantizar a diferença entre um texto hipótese (transcrições automatizadas) e o texto de referência (GT de transcrições manuais) são consideradas a quantidade de substituições, exclusões e inserções necessárias para tornar o texto de hipótese igual ao GT (MORRIS; MAIER; GREEN, 2004). A Equação 1 explicita esta relação:

$$WER = \frac{S + \mathrm{E} + I}{N} \quad (1)$$

Em que:

WER = taxa de erro de palavras;

S = quantidade de palavras substituídas no texto hipótese;

E = quantidade de palavras excluídas no texto hipótese;

I = quantidade de palavras inseridas no texto hipótese;

N = quantidade de palavras no GT.

A Tabela 1 explicita o cálculo do WER para o exemplo de um áudio real, contabilizando a quantidade de palavras substituídas (S = 9), excluídas (E = 3), inseridas (I = 6) em razão do total de palavras do GT (N = 27), resultando, de acordo com a Equação 1, em WER = 0,59.

**Tabela 1 - Exemplo de cálculo do WER**

| Transcrição GT | Transcrição Hipótese | S | E | I | N |
|---|---|---|---|---|---|
| COS, ENGIE, Mickey | COS, ENGIE, Mickey **só no ano** | 0 | 0 | 3 | 3 |
| Ô Mickey, Donald Regional Sul | O Mickey, Donald Regional Sul | 0 | 0 | 0 | 5 |
| Fala Donald | **Máquina COS** | 2 | 0 | 0 | 2 |
| Converter um grupo **para** compensador **em** Salto Santiago | **Converteu o** grupo compensador Salto Santiago | 2 | 2 | 0 | 8 |
| Ok, compensar uma máquina em Salto **Santiago** | **Por que compensam a** máquina em **soluções?** | 4 | 1 | 1 | 7 |
| Isso, obrigado | **está** obrigado. | 1 | 0 | 0 | 2 |
| | | 9 | 3 | 4 | 27 |

Fonte: Elaboração própria (2021).

O GT tem o menor WER possível, ou seja, zero. Quanto maior este número, maior a distância entre o texto hipótese e a referência. Os valores aceitáveis para WER dependem do contexto de cada áudio.

Para a geração de legendas de vídeos em tempo real, um dos cenários mais delicados em termos de WER, são identificados como complicadores a qualidade

de captação do áudio, a extensão do vocabulário e a flexibilidade no estilo da fala, que pode variar de monólogos, diálogos, cenas com efeitos sonoros ou músicas. Nos cenários mais críticos, podem ser considerados aceitáveis WER da ordem de 0,5. No extremo oposto, uma das aplicações mais bem avaliadas em termos de WER são os assistentes de voz. Neste cenário a qualidade da captação de áudio é alta, tipicamente 16 kHz, o vocabulário é estreito e orientado a alguns comandos e palavras-chave e a pronúncia tende a ser clara e pausada. Nestes casos, são usuais WER da ordem de 0,1 (MICROSOFT, 2021).

Embora não haja uma classificação para centros de operação, sobretudo para o setor elétrico, o cenário que mais se aproxima são as centrais de atendimento telefônico*.* Neste cenário em que a qualidade de áudio é baixa, geralmente 8kHz, existe a participação de dois humanos utilizando discurso relaxado e há possibilidade de sobreposição entre eles. Além disso, o vocabulário é estreito e específico de um domínio, o estilo da comunicação é relativamente estruturado, com existência de uma fraseologia, mas com possibilidades de flexibilização. Ainda, a presença de outras conversas em plano de fundo pode aumentar a quantidade de erros por Inserção. Como complicador, centros de operação em ambientes industriais estão sujeitos a ruídos de fundo que podem sobrepor os interlocutores e confundir os modelos, gerando erros de deleção. Por fim, a utilização constante de nomes próprios, como identificação de operadores e ativos, pode causar erros de substituição. Nestes ambientes são aceitáveis WER de até 0,3 (MICROSOFT, 2021).

Para cálculo do WER foi utilizada a biblioteca *Levenshtein,* do *Python*, especializada em calcular a quantidade de edições, segregadas em substituições, inserções e exclusões, necessárias para que um texto de hipótese seja convertido em um texto de referência (PYTHON-LEVENSHTEIN, 2021). Com estas informações, foram calculados, além do WER, os indicadores *Match Error Rate* (MER), *Word Information Lost* (WIL) e *Word Information Preserved* (WIP). MER é semelhante ao WER, mas normaliza os valores entre zero e um e embora utilize uma escala mais amigável, é pouco utilizada na literatura, dificultando quaisquer comparações. WIL e WIP demandam um processamento maior, e embora funcionem bem para uma amostra limitada de áudios e transcrições (200, no caso deste trabalho), tendem a ser ineficientes ao lidar com grandes quantidades de dados, limitando a adoção da metodologia aqui proposta para cenários maiores. Devido à natureza deste trabalho,

optou-se por utilizar apenas a métrica WER (MORRIS; MAIER; GREEN, 2004). O algoritmo utilizado para o cálculo das métricas é exibido no APÊNDICE B.

Para o tratamento dos textos das transcrições foi utilizada a biblioteca *jiwer*, do *Python*, que possui uma gama de funções de pré-processamento que auxiliam a manipulação de textos e transcrições, como a conversão de todo o conteúdo para letras minúsculas ou maiúsculas, remoção de espaços múltiplos, conversão de frases para listas de palavras, remoção de palavras específicas, remoção de caracteres especiais como tabulações e quebras de linhas, remoção de pontuações e substituir palavras por sinônimos (JIWER, 2020). O APÊNDICE C exibe os algoritmos utilizados para manipulação das transcrições.

### 2.3.2 WER por componente

Um aspecto que pode ser encarado como uma limitação do WER é o fato de sumarizar a qualidade de uma transcrição a um indicador único. Na Equação 1 é possível perceber que o WER considera a quantidade de palavras excluídas, substituídas e incorretas, mas não detalha cada uma das componentes. Para muitas aplicações, reduzir a complexidade inerente da comparação da qualidade de transcrições a uma análise unidimensional é justamente o que motiva a utilização do WER. Para o propósito deste trabalho, entretanto, de criação de um vocabulário fonético que otimize as transcrições do setor elétrico, além de definir a taxa de erros, é interessante conhecer com profundidade quais destes erros são mais relevantes (POPOVIC; NEY, 2007). Nas Equações 2, 3 e 4 são propostas três métricas complementares:

$$WER_S = \frac{S}{S + E + I} \quad (2)$$

$$WER_E = \frac{E}{S + D + I} \quad (3)$$

$$WER_I = \frac{I}{S + E + I} \quad (4)$$

Em que:

$WER_S$ = taxa de erros de palavras substituídas;

$WER_E$ = taxa de erros de palavras excluídas;

$WER_S$ = taxa de erros de palavras inseridas;

S = quantidade de palavras substituídas no texto hipótese;

E = quantidade de palavras excluídas no texto hipótese;

I = quantidade de palavras inseridas no texto hipótese.

Esta análise tridimensional do WER auxilia a entender os pontos fortes e fracos de cada solução estudada e serve como direcionador dos esforços para melhoria da qualidade das transcrições. O algoritmo para cálculo das componentes do WER é exibido no APÊNDICE B.

### 2.3.3 *Keyword Error Rate* (KER)

Apesar do WER ser um indicador amplamente utilizado para a validação de modelos *StT,* a métrica considera igualmente todas as palavras transcritas, não importando a sua relevância na informação transmitida (KLAKOW; PETERS, 2002). Assim, para fins de cálculo da WER, palavras genéricas como artigos definidos ou indefinidos, pronomes e conectores, são tão significativos como termos técnicos específicos, embora esta segunda categoria agregue mais valor à informação. Então pode haver situações em que um valor de WER alto não corresponda à capacidade do sistema de transcrição de extrair informações relevantes. Pensando em contornar esta limitação, existe a métrica KER, que computa os erros e faltas de palavras-chave no texto de hipótese, em comparação com o GT. A Equação 5 exibe esta relação:

$$KER = \frac{E + F}{N} \quad (5)$$

Em que:

KER: taxa de erro de palavras-chave;

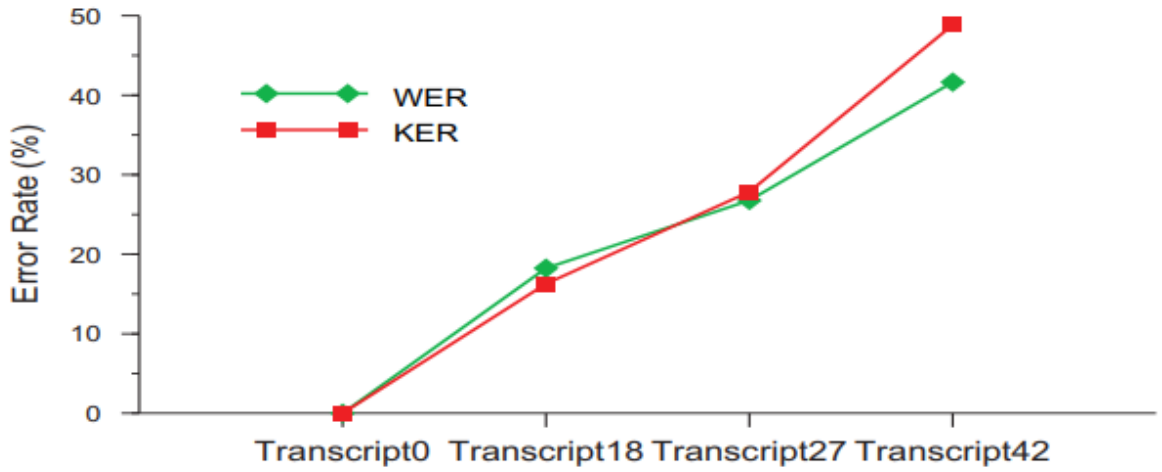
E: quantidade de palavras-chave erradas no texto de hipótese, ou falso-positivas;

F: quantidade de palavras-chave faltantes no texto de hipótese;

N: quantidade de palavras-chave no texto de referência.

Em (PARK, *et al.*, 2008) é realizada uma comparação entre as métricas WER e KER. A conclusão é que a KER, embora possa apresentar resultados melhores, é mais volátil que a WER. No referido trabalho foram comparados o GT a três conjuntos de transcrição e pode-se inferir que para valores baixos de WER (0,1824), o KER (0,1628) apresenta melhores resultados. Para valores médios, WER (0,2677) e KER (0,2785) têm desempenhos similares. Para valores altos de WER (0,4164), como o objeto de estudo desta dissertação, o KER (0,4884) tende a apresentar desempenhos inferiores. A Figura 3 exibe os valores de WER e KER para cada um dos conjuntos de transcrição.

**Figura 3 - Comparação de WER e KER**




Fonte : PARK, *et al.* (2008)

Esta conclusão, entretanto, não desencoraja a utilização do KER, pois a abordagem utilizada neste trabalho se destaca da proposta em (PARK, *et al.*, 2008) por três pontos:

a) Definição das palavras-chave: em (PARK, *et al.,* 2008) é proposto um método estatístico para definição das palavras-chave, que consiste em calcular a razão das probabilidades de uma palavra estar presente em uma transcrição do domínio específico e em transcrições genéricas. Termos cujas probabilidades em textos específicos sejam muito superiores às probabilidades em textos genéricos são considerados palavras-chave. Embora este método permita elencar palavras-chave sem a necessidade de supervisão de um especialista, ele teve como saída alguns vícios de linguagem ("*um*", "*yeah*", "*ah*") e palavras que embora pertençam ao domínio de estudo, agregam pouca informação ("*click*" ou "*server*", no contexto de uma central de ajuda de informática). As palavras-chave desta dissertação foram elencadas manualmente pelo autor, tomando como base a experiência empírica nas atividades relacionadas ao centro de operação, bem como sua familiaridade com os conteúdos dos áudios, jargões técnicos e nomes próprios dos ativos. Uma possível contribuição de trabalhos futuros seria a criação de listas de palavras-chave mais abrangentes. A lista completa de palavras-chave escolhidas pode ser consultada na Tabela 2. Embora menos replicável para contextos diferentes, esta abordagem evita discrepâncias, como o fato de as cinco palavras-chave mais bem ranqueadas em (PARK, *et al.*, 2008) serem *stop words* (definição a seguir);

**Tabela 2 - Lista de Palavras-Chave**

| | | | |
|---|---|---|---|
| ONS | Partida a Frio | Subcorrente | Aerogerador |
| IHM | Tomada d'Agua | Subfrequência | Cana Brava |
| ENGIE | Fuga Terra | Montante | Estreito |
| PGI | Comando | Jusante | Itá |
| SGI | Evento | Compensador | Jaguara |
| ANEEL | Proteção | Perturbação | Machadinho |
| CNOS | Transformador | Sistêmica | Miranda |
| COSR | Geração | Umburanas | Passo Fundo |
| COS | Programação | Black start | Ponte de Pedra |
| COG | Válvula | Termelétrica | São Salvador |
| UHE | Partida | Para raio | Salto Osório |
| PCH | Tubulação | Disjuntor | Salto Santiago |
| MUST | Indisponibilidade | Energizada | Lacerda A |
| Partida a Quente | Comporta | Normalizada | Lacerda B |
| Sincronismo | Seccionadora | Restrição | Lacerda C |
| Oscilação | Inversor | Turbina | José Gelázio |
| Rotação | Potência Ativa | Caldeira | Rondonópolis |
| Turbina | Potência Reativa | Precipitador | Cacimbas |
| Instabilidade | Fator de Potência | Vertimento | Estrela |
| Manutenção | Manobra | Vertedouro | Ouro Verde |
| Alarme | Retificador | Deplecionar | Santa Mônica |
| Gerador | Sobretemperatura | Amperes | Flecheiras |
| Falha | Sobretensão | Volts | Guajirú |
| Subtensão | Sobrecorrente | Watts | Mundaú |
| Serviço Auxiliar | Sobrefrequência | Hidrelétrica | Trairi |
| Trafo | Sobrevelocidade | Parque Eólico | Campo Largo |

Fonte: Elaboração própria (2021).

Para o cálculo do KER o primeiro passo consiste em realizar uma busca pelos termos da Tabela 2 nos texto de hipótese e GT. Em seguida, são comparadas as listas e verificadas as palavras-chave que estão no texto de hipótese e não no GT (erros) e as palavras-chave que estão no GT, mas não no texto de hipótese (faltas). De posse dessas informações é aplicada a Equação 5. O APÊNDICE D exibe os algoritmos de cálculo do KER.

b) Remoção de *stop words*: ao analisar a ocorrência das palavras em um texto, a primeira suposição pode ser de que os termos mais recorrentes sejam os mais relevantes. Esta intuição, entretanto, é o oposto da realidade. As alternativas mais frequentes em qualquer documento serão palavras comuns, como “o”, “a”, “e”, entre

outras que ajudam a construir ideias, mas que não possuem significados isoladamente. Por vezes a relevância de algumas palavras pode variar dependendo do contexto analisado, mas é usual filtrar as palavras mais comuns de um idioma, ou *stop words*, durante o pré-processamento de linguagem natural (RAJARAMAN; ULLMAN, 2011). Nesta dissertação, antes de se analisar o KER, foi implementado um filtro de *stop words* do idioma português. O objetivo é remover pronomes, preposições, conjunções e artigos que tem baixa relevância na informação transmitida pelo texto e evitar resultados como em (PARK, *et al.*, 2008), no qual dentre as 30 palavras-chave elencadas, estão presentes termos como "*wanna*", "*gotta*" e "*thank*", exemplos de *stop words* do inglês (idioma original dos áudios analisados).

c) *Stemming:* é o processo de aglutinação das variações de uma palavra, que têm o mesmo significado, em uma representação única, mesmo que esta representação não tenha significado no idioma original. Esse processo, embora desenvolvido originalmente para o inglês e posteriormente adaptado para outros idiomas, encontra bastante aplicação no português, em que um conceito ou radical pode ser representado por muitas palavras (ALVARES; GARCIA; FERRAZ, 2005). Esta estratégia desempenha um papel importante na quantificação da qualidade das transcrições, visto que o WER considera como substituições as palavras transcritas com grau, gênero ou número diferentes do original, apesar de a informação em si ser preservada. A aplicação de técnicas de *steamming* no pré-processamento das transcrições é outro item que complementa os estudos realizados em (PARK, *et al.*, 2008). É necessário, contudo, atentar para os dois tipos de erros que podem acontecer no processo:

   i. *over stemming:* quando há excesso de remoção de letras, de modo que palavras com diferentes significados sejam mescladas em um mesmo radical. A Tabela 3 exemplifica este erro.

   ii. *under stemming:* quando são preservadas muitas letras, de modo que palavras com o mesmo significado são agrupadas em diferentes radicais. A Tabela 4 exemplifica este erro.

**Tabela 3 - Exemplo de *over stemming***

| Significado | Palavra | Radical |
|---|---|---|
| compensar | compensador | comp |
| comparar | comparação | comp |

Fonte: Alvares; Garcia; Ferraz (2005).

**Tabela 4 - Exemplo de *under stemming***

| Significado | Palavra | Radical |
|---|---|---|
| movimentar | movimentação | movim |
| movimentar | movido | movid |

Fonte: Alvares; Garcia; Ferraz (2005).

Para as etapas de remoção de *stop words* e *stemming* foi utilizada a biblioteca *Natural Language Toolkit* (NLTK), do *Python*. Essa biblioteca possui ferramentas léxicas para vários idiomas, incluindo opções em português, como o Removedor de Sufixos da Língua Portuguesa (URGS, 2010), *Portuguese Treebank* (AFONSO, *et al.*, 2002) e o *MAC-MORPHO: Brazilian Portuguese news text with part-of-speech tags* (USP, 2015). Ela possui ainda um conjunto de funções para manipulação de dados de NLP em grande escala (NLTK PROJECT, 2020). O APÊNDICE E e APÊNDICE F apresentam os algoritmos de remoção de *stop words* e aplicação de *stemming*, respectivamente.

### 2.3.4 *Named Entity Recognition* (NER)

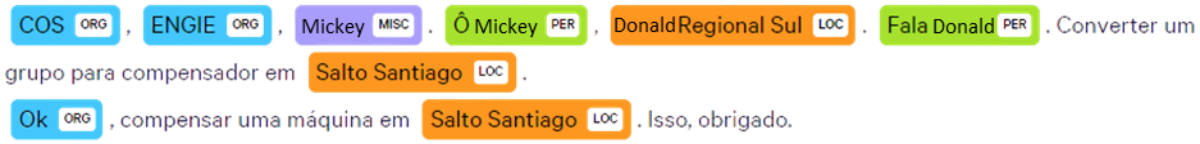
É consenso no meio de estudos de NLP que a quantidade de palavras corretas é um bom indicador para avaliação da qualidade das transcrições (NLTK PROJECT, 2020). Mais importante do que minimizar o WER, entretanto, é o fato dos modelos de transcrição serem capazes de extrair a compreensão do objeto de estudo da linguagem falada (ZECHNER; WAIBEL, 2000). Desta forma, metodologias de avaliação capazes de classificar, reconhecer e comparar estruturas relevantes nos textos de GT e hipótese podem ser usadas como indicadores de assertividade da compreensão, independente da precisão da transcrição de cada palavra (WANG; ACERO; CHELBA, 2003).

Objetivando propor uma métrica de assertividade da compreensão, neste trabalho foi utilizado o NER. Trata-se de uma técnica amplamente estudada e utilizada em tarefas de NLP. Originalmente concebida para extração de informações

estruturadas a partir de textos não estruturados, o NER consiste em identificar unidades de informação e classificá-las em entidades, como nomes de pessoas, organizações, locais e expressões numéricas, como horários, datas, valores monetários e percentuais (NADEAU; SEKINE, 2007).

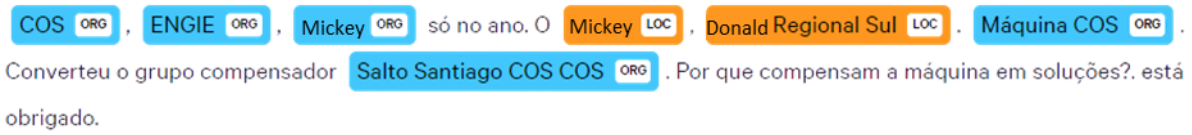
Estas características podem ser usadas para identificação de entidades específicas do contexto de centros de operação do setor elétrico e comparar os dados extraídos das transcrições de hipótese e do GT. Neste trabalho foram elencadas três entidades: pessoa (PER); organização (ORG) e localização (LOC). As Figuras 4 e 5 exibem os resultados do reconhecimento das entidades que o modelo NER encontrou em uma transcrição de referência e em uma transcrição candidata, respectivamente, de um dos áudios reais do COS da ENGIE.

**Figura 4 - Exemplo de NER - transcrição de referência**




Fonte: Elaboração própria (2021).

**Figura 5 - Exemplo de NER - transcrição de hipótese**




Fonte: Elaboração própria (2021).

Nas Figuras 4 e 5 é possível observar que a classificação do NER não é completamente precisa. Na Figura 4, por exemplo, o termo “Ok” foi classificado como ORG. Isso acontece pois o modelo do NER também precisa ser treinado e ajustado. Para aplicações de classificação de chamadas, por exemplo, um modelo aderente é fundamental para resultados confiáveis. Para os fins de comparação de transcrições, entretanto, um modelo moderadamente funcional pode ser empregado. Isto porque ao aplicar o mesmo critério NER às transcrições GT e hipótese, é possível evidenciar as diferenças entre as transcrições, sendo esperados que os erros de classificação, quando existam, sejam igualmente aplicados a ambos os textos.

Para quantificar as divergências ou semelhanças entre o NER aplicado a dois textos, neste trabalho são propostos os indicadores de razão da ocorrência de entidades gerais e por tipo, reconhecidas no texto GT e nas transcrições hipóteses, como explicitado nas Equações 6 e 7.

$$ER = \frac{E_{GT} - E_{hip}}{E_{GT} + E_{hip}} \tag{6}$$

$$ER_{ent} = \frac{E_{GT_{ent}} - E_{hip_{ent}}}{E_{GT} + E_{hip}} \tag{7}$$

Em que:

ER: razão das entidades reconhecidas nos textos de hipótese e GT;

$E_{hip}$: quantidade de entidades reconhecidas no texto de hipótese;

$E_{GT}$: quantidade de entidades reconhecidas no texto de GT;

ent: entidades pré-definidas (PER, ORG ou LOC).

As razões das entidades reconhecidas são valores que podem variar de 1 a -1. Quanto mais próximo de zero, mais preciso o reconhecimento, de sorte que um valor zero significa que foram reconhecidas a mesma quantidade de entidades no GT e na hipótese. Valores positivos indicam que algumas entidades do GT não foram reconhecidas na hipótese, ao passo que valores negativos indicam a presença de falso positivos, ou seja, que a hipótese tem mais entidades reconhecidas que o GT. A Tabela 5 exibe a aplicação das Equações 6 e 7 aos reconhecimentos de entidades das Figuras 4 e 5.

**Tabela 5 - Exemplo de reconhecimento de entidades utilizando NER**

| | | | | |
|---|---|---|---|---|
| $E_{hyp} = 4$ | $E_{hyp_{ORG}} = 0$ | $E_{hyp_{PER}} = 0$ | $E_{hyp_{LOC}} = 2$ | $E_{hyp_{MISC}} = 0$ |
| $E_{GT} = 7$ | $E_{GT_{ORG}} = 2$ | $E_{GT_{PER}} = 2$ | $E_{GT_{LOC}} = 3$ | $E_{GT_{MISC}} = 1$ |
| $ER = 0{,}27$ | $ER_{ORG} = 1$ | $ER_{PER} = 1$ | $ER_{LOC} = 0{,}2$ | $ER_{MISC} = 1$ |

Fonte: Elaboração própria (2021).

Para a análise das transcrições de hipótese e GT e consequente cálculo do NER foi utilizada a biblioteca *spaCy*, do *Python*. A biblioteca apresenta um bom

desempenho ao lidar com grandes quantidades de dados, com suporte a 64 idiomas e possui funcionalidades relacionadas a NLP como componentes para NER, rotulação de fala; segmentação de sentenças; classificação textual entre outros facilitadores (SPACY, 2021). O APÊNDICE G exibe os algoritmos utilizados para cálculo do NER e exibição dos resultados.

### 2.3.5 Outras métricas e estratégias

Como demonstrado neste capítulo, a avaliação e análise dos erros de transcrição é uma tarefa importante e complexa. A literatura é ampla no que tange técnicas, métricas e abordagens para definições das qualidades das transcrições automáticas. Somada a isto, existe a possibilidade de adaptação das técnicas e indicadores utilizados em avaliação de traduções automáticas (PAPINENI, *et al.*, 2002).

Como um dos objetivos deste trabalho é concentrar as informações relevantes para utilização de transcrições de áudios do setor elétrico, esta seção é dedicada a elencar outros indicadores que podem ser utilizados como referência de estudos vindouros:

a) *Match error rate* (MER): WER não é um indicador normalizado. No caso hipotético de uma determinada transcrição substituir cada uma das palavras do texto de referência, segundo a Equação 1, o WER seria 1. Caso além da substituição, para cada palavra do GT fosse inserida uma nova palavra, o WER seria 2. Ambas as transcrições não acertaram nenhuma palavra e ainda assim a análise desses números pode sugerir uma diferença considerável de performance entre elas. O MER busca normalizar o WER entre os valores 0, para uma transcrição exata, e 1 para uma hipótese sem nenhum acerto. Além de mais intuitivo, o MER representa a probabilidade de uma palavra ser corretamente transcrita, dado um determinado modelo (MORRIS; MAIER; GREEN, 2004).
b) *Position independent word error rate* (PER): semelhante ao WER, mas compara as palavras do GT e do texto candidato, independentemente da sua posição. Pode evitar a computação de erros indevidos, mas pode camuflar erros nos casos em que a ordem seja importante (POPOVIC; NEY, 2007). No

contexto de chamadas dos centros de operação do setor elétrico, nas quais são exigidas as confirmações das informações (ONS, 2020) e, portanto, há muitas repetições de palavras, pode desconsiderar alguns erros de procedimentos.

c) *Part of speech (*POS): decompõe os resultados de WER e PER em classes de interesse, como classes gramaticais (verbos, advérbios, substantivos entre outras) das palavras erroneamente transcritas. Permite comparar modelos de transcrição sob uma ótica distinta, bem como identificar os pontos mais sensíveis de ajustes (POPOVIC; NEY, 2007).
d) *Bilingual evaluation understudy* (BLEU): usa uma média ponderada de correspondência em frases de comprimento variável, entre o texto de referência e de hipótese (PAPINENI, *et al.*, 2002).
e) *National Institute of Standards and Technology* (NIST) *Score*: possui uma base de dados de texto e utiliza *n-gram* para definir a probabilidade da próxima palavra, considerando as *n* palavras precedentes (DODDINGTON, 2002).
f) *Metric for evaluation of translation with explicit ordering* (METEOR): conta o número de correspondências exatas entre o GT e a hipótese. Posteriormente, as palavras sem correspondência direta são convertidas em radicais (*stemming*) ou sinônimos. Por fim, é realizada uma nova comparação (BANERJEE; LAVIE, 2005).
g) *Translation error rate* (TER): contabiliza a quantidade de correções necessárias que um humano precisaria fazer em um texto candidato, para que ele correspondesse exatamente ao GT (SNOVER, *et al.*, 2006).
h) *Cover disjoint error rate* (CDER): baseado na distância entre os textos hipóteses e de referência, como o WER, mas permite reordenar os blocos de informação (LEUSCH; UEFFING; NEY, 2006).
i) *Relative information lost* (RIL): indicador dual ao RIP. Ele fornece uma medida estatística da dependência mútua entre o texto de GT e a hipótese, considerando conjuntos não ordenados de pares de palavras do GT e do texto candidato. Não é tão utilizado devido ao alto custo computacional para sua obtenção e devido à forte dependência do contexto, sendo que qualquer comparação, palavra a palavra, resultaria em zero erros (MORRIS; MAIER; GREEN, 2004).

j) *Relative information preserved* (RIP): indicador dual ao RIL. Ele fornece uma medida estatística da independência entre o texto de referência e a hipótese, considerando conjuntos não ordenados de pares de palavras do GT e do texto candidato. Não é tão utilizado devido ao alto custo computacional para sua obtenção e devido à forte dependência do contexto, sendo que qualquer comparação, palavra a palavra, resultaria em zero erros (MORRIS; MAIER; GREEN, 2004).
k) *Word information lost* (WIL): evolução do RIL, que alcança resultados muito próximos, sem os inconvenientes do alto custo computacional e pode ser utilizado em análises um a um. (MORRIS; MAIER; GREEN, 2004).
l) *Word information preserved* (WIP): evolução do RIP, que alcança resultados muito próximos, sem os inconvenientes do alto custo computacional e pode ser utilizado em análises um a um. (MORRIS; MAIER; GREEN, 2004).

## 2.4 Tratamento de arquivos de áudio

Durante os primeiros testes de transcrição foi identificado que os modelos não estavam desempenhando bem, apresentando WER médio de 0,65. Objetivando minimizar o WER médio, foram utilizadas técnicas de tratamento dos arquivos de áudio e foi possível observar que o tratamento, embora elementar, apresentou melhorias substanciais na qualidade das transcrições.

O aprofundamento das técnicas e teorias de tratamento de áudio não fazem parte do escopo deste trabalho. Assim, esta seção descreve algumas propostas, considerando a simplicidade de implementação e a melhoria da qualidade das transcrições.

Os áudios são gravados entre centros de operação dos agentes e do ONS, entre os centros de operação do agente e suas instalações ou, para os agentes que não contam com centros de operação, entre o ONS e suas instalações (ONS, 2020). Do ponto de vista de qualidade das gravações de áudio, os centros de controle utilizam um vocabulário relativamente restrito e uma estrutura fixa, com uma proposta de fraseologia padronizada (ONS, 2020). São utilizados muitos nomes próprios, como nomes de operadores, ativos e cidades e por vezes há sobreposição entre os

interlocutores e conversas de fundo. Todas essas características são compartilhadas por centrais de atendimento telefônico, ambientes nos quais são esperados WER da ordem de 0,3 (MICROSOFT, 2021). Como agravante, os áudios dos interlocutores nas instalações locais estão sujeitos à ruídos inerentes à atividade industrial e, apesar da fraseologia, o discurso relaxado foi identificado com bastante regularidade.

Estas características motivaram o estudo e a aplicação de ferramentas de tratamento e manipulação de áudios. Nas últimas décadas, métodos computacionais foram desenvolvidos para estimar a qualidade perceptual do áudio. Baseado no trabalho de (TORCOLI; KASTNER; HERRE, 2021), foram aqui propostos os seguintes melhoramentos:

a) amplificação do sinal: objetivando garantir que todos os arquivos fossem audíveis, por pessoas e modelos, foram aplicados ganhos de 2dB;
b) normalização: garante homogeneidade na amplitude dos sinais de todos os interlocutores;
c) filtros passa-alta: filtra ruídos de frequências inferiores a 120 Hz;
d) filtros passa-baixa: filtra ruídos de frequências superiores a 4 kHz.

Em adição aos tratamentos de áudio, foi especificada uma alteração nos parâmetros do gravador de chamadas para que os arquivos sejam armazenados com taxas de 16 kHz, em oposição ao padrão utilizado, de 8 kHz. Embora os arquivos salvos com uma taxa de amostragem maior demandem mais espaço de armazenamento, este valor não é significativo. A qualidade superior dos áudios é refletida no aumento da qualidade das transcrições (MICROSOFT, 2021).

Os áudios são gravados originalmente no formato estéreo, em que cada um dos ramais é salvo em uma faixa (um interlocutor é reproduzido no canal esquerdo, e o outro no direito). A fim de tornar os áudios compatíveis com as diferentes ferramentas de transcrição, foi utilizado o Audacity (2021). Este *software*, gratuito e de código aberto, é dedicado à edição de áudio e possui funcionalidade para converter os arquivos para o formato mono, onde ambos os sinais são reproduzidos em uma mesma faixa.

Após a aplicação destas medidas, o WER médio passou para 0,47, uma melhoria de quase 28%.

Para a manipulação dos arquivos de áudio foi utiliza a biblioteca *pySox*, do *Python*. A biblioteca foi capaz de aplicar as manipulações a todos os áudios em 17 segundos, enquanto o *pydub*, outra biblioteca consolidada do *Python*, precisou de três minutos. A *pySox* possui uma gama de funções relacionadas a tratamento de áudio, como: a extração de informações dos arquivos; a concatenação de áudios; a aplicação de ganhos; normalização, *fade-in* e *fade-out*; a alteração do número de canais, entre outros funcionalidades (PYSOX, 2016).

## *2.5* Definição do *ground truth*

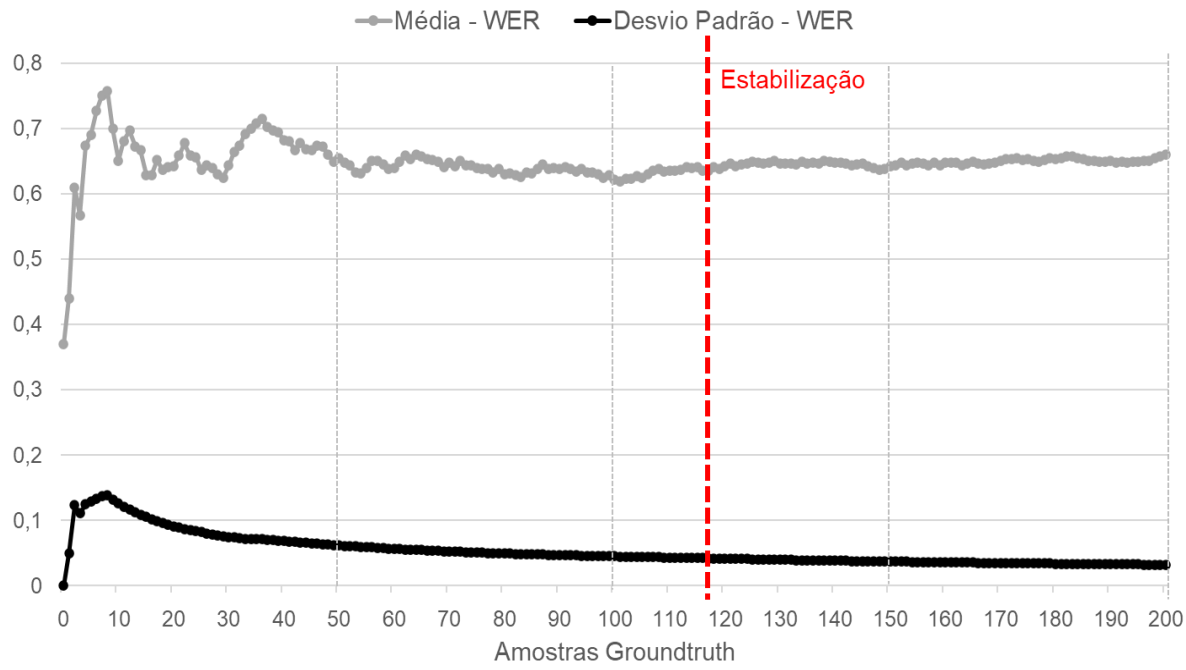
Em projetos relacionados à aplicação de ciência de dados é necessária a definição da quantidade mínima de amostras que possibilite a realização de experimentos em um ambiente que represente, de maneira satisfatória, a realidade (STROBEL; CLEMATIDE, 2020). Esta seção tem por objetivo propor uma abordagem para definir o tamanho da amostra de transcrição manual das chamadas.

Para tanto, foi utilizada a métrica WER. Após analisar o WER de várias transcrições, é esperado que haja uma média e um desvio padrão dos WER de todas chamadas. Para as primeiras amostras, é provável que o desvio padrão seja elevado. Á medida que são inseridas amostras ao GT é esperado que haja uma convergência na média e uma diminuição no desvio padrão, até o patamar em que a adição de mais amostras, pouco interfira com as métricas. Este é considerado o ponto de estabilidade do sistema (STROBEL; CLEMATIDE, 2020).

Inicialmente foi utilizado um GT de transcrições manuais realizadas pelo autor, composto por 50 amostras. Ao verificar a média do WER, observou-se que ela apresentava variação na tendência, enquanto o desvio padrão exibia uma tendência de baixa, sugerindo que novas amostras eram necessárias. Decidiu-se adicionar mais 50 amostras ao GT e os mesmos resultados, embora suavizados, foram encontrados. Repetiu-se o processo, totalizando 150 amostras. Neste momento, foi identificada uma estabilização da média e do desvio padrão, pouco antes das 120 amostras. A fim de assegurar que o comportamento se manteria, foram transcritos mais 50 áudios e o

comportamento de convergência se confirmou, totalizando as 200 transcrições que foram utilizadas nos estudos e desenvolvimentos deste trabalho. A Figura 6 exibe a variação da média e desvio padrão do WER, em função da quantidade de transcrições manuais.

**Figura 6 - Média e desvio padrão do WER pelo número de transcrições**




Fonte: Elaboração própria (2021).

É possível observar na Figura 6 que os valores estabilizam pouco antes das 120 amostras. As transcrições seguintes pouco influenciam nos valores da média e do desvio padrão de WER. Desta forma, para o trabalho proposto, 120 amostras seria o valor mínimo necessário para representar o cenário real. Pensando em garantir uma massa de dados condizente e contar com uma margem de segurança, foram transcritos 200 chamadas manualmente.

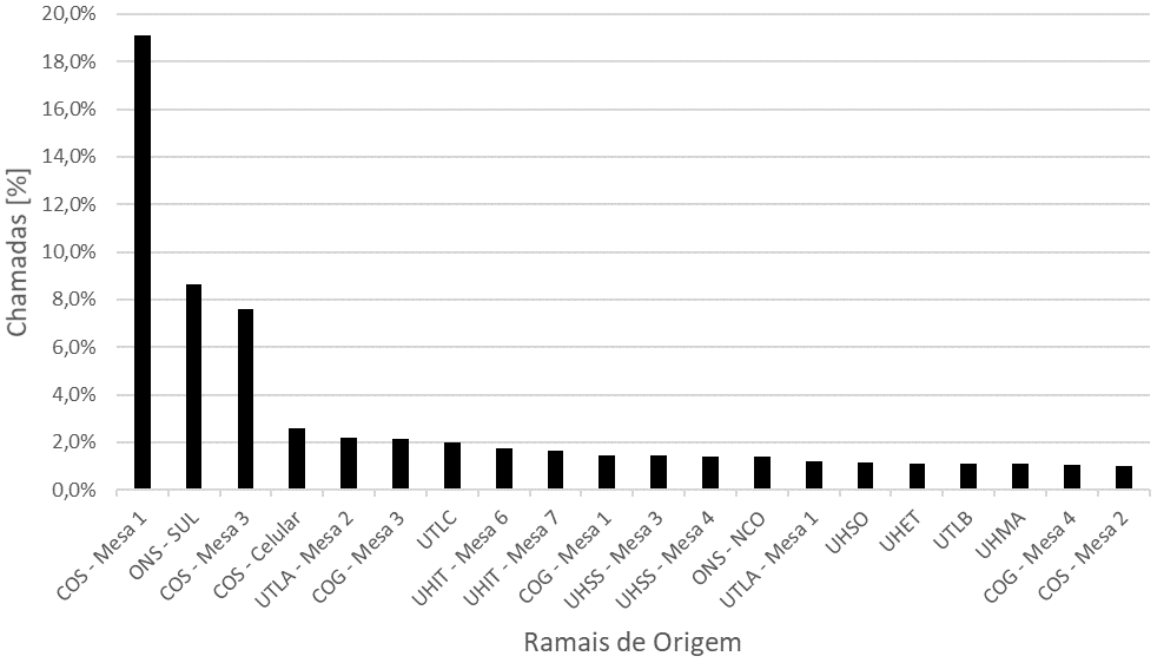
### 2.5.1 Análise dos metadados

Além da quantidade das chamadas necessárias é importante analisar a qualidade das amostras. Com o objetivo de escolher áudios condizentes, foi realizada uma análise exploratória dos metadados dos mais de 550 mil arquivos. Além das informações intrínsecas às chamadas de áudio, cada arquivo conta com alguns metadados, como: ramal de origem; ramal de destino; instante da ligação e duração

da chamada. Com base nestas informações, foi possível realizar algumas inferências que serviram de critério para a escolha das 200 amostras de chamadas a serem manualmente transcritas.

No tocante aos ramais de origem, foram identificados 4335 ramais únicos e os 20 ramais que concentram mais ligações, totalizam 61% de todas as ligações. Ou seja, 0,5% de todos os ramais de origem, são responsáveis por mais da metade das ligações. Além disso, destes 20 ramais, muitos são apenas mesas distintas de um mesmo centro de operação ou usina, reforçando o caráter concentrado das chamadas. As chamadas iniciadas por estes ramais foram escolhidas como amostras de áudio a comporem o GT. A Figura 7 exibe o percentual de ligações dos 20 ramais de origem mais recorrentes.

**Figura 7 - Histograma das ligações por ramais de origem**

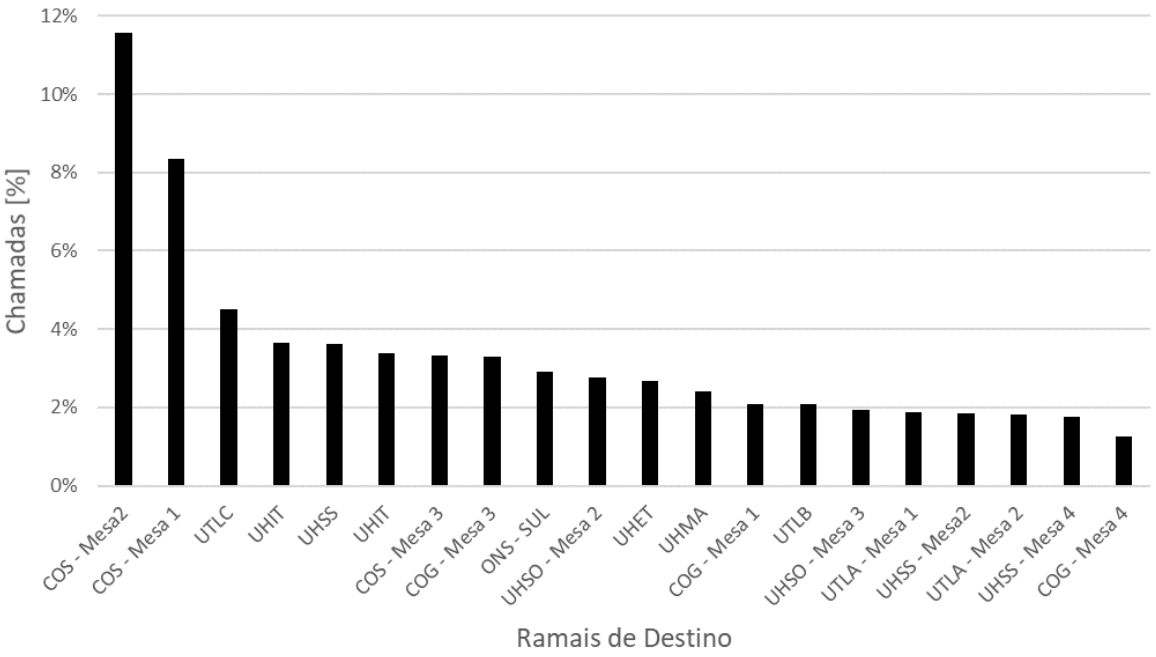



Fonte: Elaboração própria (2021).

No que tange os ramais de destino, no universo de 550 mil chamadas foram identificados 3786 ramais distintos. Apesar disso, os 20 ramais que concentram mais ligações, acumulam um total de 67% das chamadas. Ou seja, 0,5% de todos os ramais de destino, recebem cerca de dois terços de todas as ligações. Também aqui foi observado que destes 20 ramais, muitos são apenas estações diferentes de um mesmo centro de operação ou usina. As chamadas para estes ramais foram

escolhidas como amostras de áudio a serem manualmente transcritas. A Figura 8 mostra o percentual de ligações dos 20 ramais de destinos mais utilizados.

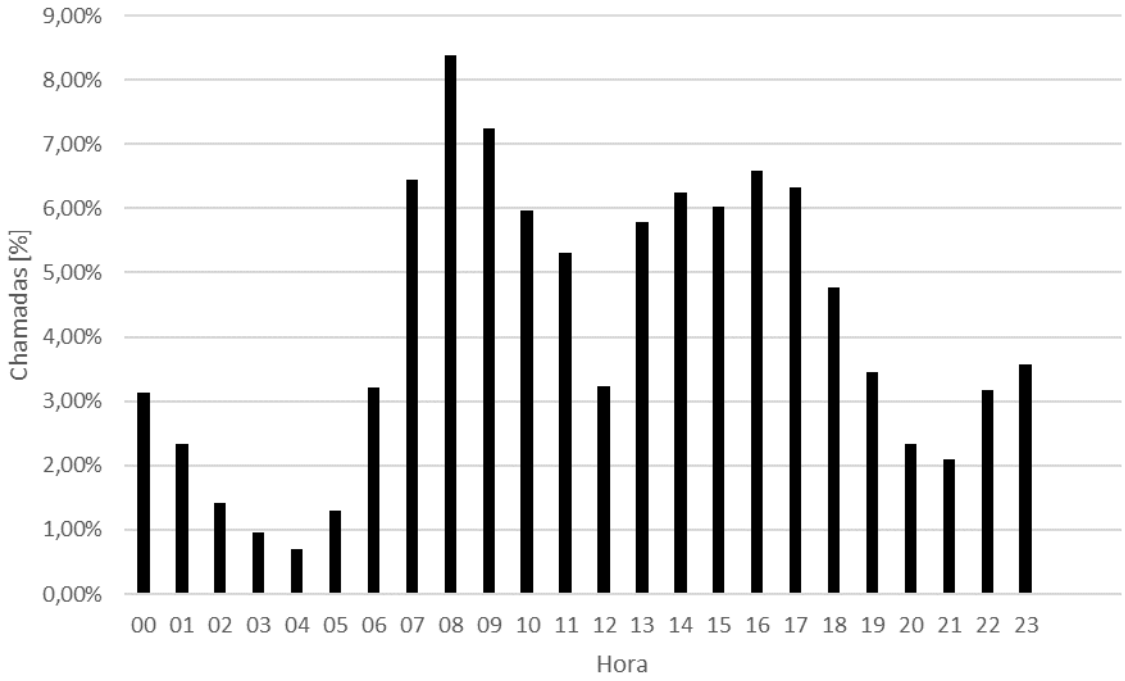
**Figura 8 - Histograma das ligações por ramais de destino**




Fonte: Elaboração própria (2021).

Sobre o horário das chamadas, é possível observar que a distribuição segue um padrão semelhante à modulação da carga ao longo de um dia típico. Embora curioso sob a ótica de dimensionamento de equipes e definição de horários de passagem de turno, no que diz respeito à escolha de áudios para criação do GT, o histograma da Figura 9 permite concluir que a melhor opção de escolhas de áudio, segundo este critério, é escolher as chamadas aleatoriamente, de sorte que as amostras reproduzam o comportamento da população geral.

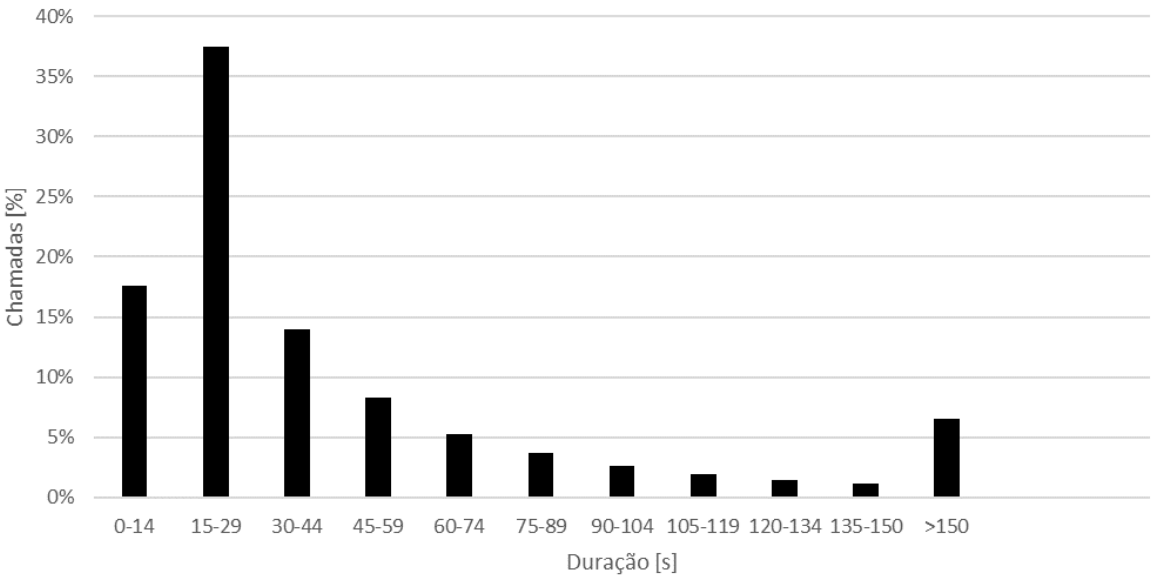
**Figura 9 - Distribuição das ligações ao longo do dia**




Fonte: Elaboração própria (2021).

No que diz respeito à duração, ao analisar o histograma de duração das ligações da Figura 10, é possível observar que embora haja áudios com duração maiores, eles representam situações particulares. As chamadas típicas dos centros de controle são curtas e diretas, sendo que 77% das ligações têm duração de até 60 segundos. Por ser uma categoria que representa com mais fidelidade as chamadas dos centros de operação, foram transcritos manualmente áudios com até um minuto de duração.

**Figura 10 - Histograma de duração das ligações**




Fonte: Elaboração própria (2021).

Por fim, após as análises dos metadados e estudo de convergência da média do WER, foi definido que a dissertação tem, como GT, uma massa de 200 transcrições, cujos ramais de origem e destino pertençam ao conjunto dos 20 ramais mais recorrentes, conforme Figuras 7 e 8. Além disso, serão utilizados apenas áudios com duração inferior a um minuto, aleatoriamente escolhidos quanto ao instante da chamada.

## 2.6 Comparação entre ferramentas de transcrição

A fim de transcrever automaticamente os áudios dos centros de operação do setor elétrico, cinco ferramentas consolidadas no mercado foram avaliadas: *Amazon Web Services* (AWS); Dígitro Retextoar; *Google Cloud Platform* (GCP); IBM Watson e *Microsoft Azure Cognitive Services* (MACS).

Trabalhar com soluções de mercado possui vantagens e desvantagens. Pode ser uma desvantagem o fato de a tecnologia ser fechada, não se ter acesso completo ao código e não se ter certeza do que ela está efetivamente fazendo. A vantagem é que a ferramenta está sendo usada por um número considerável de clientes e, também por isso, ela é constantemente corrigida e atualizada. As cinco ferramentas citadas foram utilizadas no formato de acesso livre e gratuito.

Nesta seção são aplicadas diferentes métricas a fim de elencar as soluções com melhor desempenho de transcrições. Além disso, é explorada a capacidade das soluções AWS e GCP de utilizarem vocabulários personalizados, mesmo em um nível superficial. De forma geral, e em específico sobre o contexto desta dissertação, faz-se os seguintes apontamentos:

a) AWS: para converter fala em texto, utiliza um processo de aprendizagem profunda chamada reconhecimento automático de fala (ASR). Uma das vantagens desta abordagem é tornar a ferramenta de transcrição adaptável a diferentes cenários, como os centros de operação do setor elétrico, em que os áudios de telefonia podem ter uma baixa qualidade de gravação, resultando em baixa qualidade da transcrição. A ferramenta pode ser usada para transcrever chamadas, gerar metadados para avaliação e criar arquivamento rastreável (AMAZON, 2021). Uma importante funcionalidade do AWS é a possibilidade de criação de listas de vocabulário e modelos de idioma personalizados. Para gerar transcrições especializadas em palavras de domínios, como nomes de locais, terminologia técnica ou nomes de indivíduos, é possível adicionar novas palavras ao vocabulário básico ou adaptar o próprio modelo de idioma (AWS, 2021). Embora a adaptação do próprio modelo ainda não esteja disponível para o idioma português, a funcionalidade de vocabulário personalizado se mostrou bastante promissora para o desafio da transcrição de áudios no setor elétrico. Nesta seção é avaliada a solução da AWS com e sem vocabulário personalizado.
b) GCP: o serviço de StT do Google tem três métodos principais para realizar o reconhecimento de fala (GOOGLE, 2021):
    I. Reconhecimento síncrono: executa o reconhecimento de voz e retorna os resultados depois que todo o áudio foi processado. Ele é limitado a ligações de até um minuto de duração.
    II. Reconhecimento assíncrono: analisa os dados de áudio e inicia uma operação de longa duração. Solicitações assíncronas são utilizadas para dados de áudio com duração de até 480 minutos.
    III. Reconhecimento de *streaming*: realiza reconhecimento em dados de áudio em tempo real. Ele oferece resultados provisórios enquanto o áudio está sendo capturado.

Para os fins desta dissertação, foi utilizado o reconhecimento assíncrono.

O GCP permite configurar parâmetros com o intuito de melhorar a qualidade das transcrições. Dentre as configurações que merecem destaque estão as 285 opções de idiomas. Para esta avaliação foram utilizados dois idiomas: Português (Brasil) - Comando e pesquisa e Português (Brasil) – Padrão.

Outra opção para melhorar os resultados, é especificando a origem do áudio. Isto permite que o GCP processe os arquivos usando um modelo treinado para reconhecer o áudio daquele tipo específico de origem. Para os fins desta comparação, foi utilizado o modelo de chamada telefônica, onde o áudio é gravado a uma taxa de amostragem de 8 kHz (GOOGLE, 2021).

Por fim, o GCP dispõe de uma funcionalidade chamada *boost*, capaz de adaptar as transcrições a um contexto específico. A otimização da adaptação de fala é um recurso opcional. O *boost* permite atribuir pesos numéricos a palavras ou frases, de acordo com a frequência com que eles devem ser reconhecidos nos dados de áudio (GOOGLE, 2021). Como o autor dispõem de experiência nos termos mais relevantes e recorrentes dos áudios, para esta comparação será utilizada a solução do GCP com e sem *boost* em palavras-chave.

c) Dígitro: empresa brasileira do setor de tecnologia, sediada em Florianópolis e desenvolvedora do Retextoar. Trata-se de uma ferramenta de transcrição idealizada para transcrever grandes volumes de áudio e vídeo, principalmente nos setores de entretenimento, publicidade e jurídico. É voltada para empresas com necessidades como dicionário de vocabulário especializado e grande volume de transcrição. Possui algoritmo desenvolvido especialmente para o português brasileiro (DÍGITRO, 2021).
d) IBM Watson: os recursos do IBM Watson para StT podem ser utilizados com arquivos de áudio e voz em tempo real. Ele é focado nos setores de serviços financeiros, assistência médica, seguro e telecomunicações Suporta reconhecimento de voz em mais de 15 idiomas e o modelo utiliza como entradas o idioma no qual o áudio é falado e a taxa em que é amostrado (IBM, 2021). Para a maioria dos idiomas o serviço suporta os modelos de banda larga e de banda estreita. Modelos de banda larga são utilizados para áudios

amostrados em 16 kHz ou acima, como aplicativos de fala em tempo real. Modelos de banda estreita são utilizados com áudios amostrados em 8 kHz, tipicamente utilizados para chamadas telefônicas (IBM, 2021). Para esta comparação foi utilizado o idioma português do Brasil e o modelo de banda estreita. Em sua versão gratuita o IBM Watson não permite a utilização de vocabulários personalizados.

e) MACS: O serviço faz parte da mesma plataforma de tecnologia de reconhecimento de voz utilizada pela Microsoft para os produtos do Office. O comportamento padrão do serviço de transcrição é usar o modelo de linguagem universal, desenvolvido para cenários de conversa e de ditado. Ao usar a conversão de fala em texto em um ambiente dedicado, é possível criar e treinar modelos de linguagem e de pronúncia personalizados. Em sua versão gratuita, o serviço da MACS oferece suporte apenas para arquivos de áudio do formato WAV, sem compatibilidade com padrões consolidados de mercado, como o MP3. O serviço suporta a conversão de vários arquivos, desde que estejam armazenados no serviço de nuvem da Microsoft. A transcrição de áudios em lotes permite a transcrição assíncrona de grandes volumes de dados de áudio e foi a opção utilizada para esta comparação (MICROSOFT, 2021).

### 2.6.1 Critérios de Comparação

Após elencar as cinco soluções de mercado, foram iniciados alguns testes comparativos. Embora neste capítulo os testes sejam apresentados de uma forma sequencial e direta, na prática, eles foram tornados mais completos de forma incremental, à medida que os resultados anteriores não eram capazes de prover, com a assertividade desejada, uma medida suficientemente precisa entre o desempenho das diferentes soluções.

Outro fator que difere da cronologia dos eventos é o fato de que nesta dissertação são avaliadas, simultaneamente, as personalizações de vocabulário da solução da AWS e GCP. Estes testes foram os últimos a serem realizados, justamente com as soluções que se mostraram mais promissoras.

Com o intuito de evitar repetições e simplificar a apresentação dos métodos de comparação, os resultados serão apresentados todos simultaneamente, totalizando sete opções.

Por fim, como foram utilizados diferentes indicadores com escalas de valores não compatíveis, optou-se por utilizar um sistema de notas, que variam de 7 para a solução com melhor desempenho, e 1 para aquela que apresentou o pior resultado. Embora esta escolha possa dificultar a visualização da distância do desempenho entre duas soluções, ela é útil no sentido de determinar se existem soluções completamente dominadas ou dominantes sob os diferentes critérios.

Os dois primeiros testes foram realizados utilizando a média do WER para as transcrições. Além disso, objetivando avaliar o desempenho das ferramentas com diferentes formatos de áudio, foram realizados os testes com arquivos WAV e MP3. As Tabelas 6 e 7 exibem os resultados:

**Tabela 6 - Comparação WER entre as soluções – Arquivos WAV**

| Soluções | WER Médio (WAV) | Nota |
|---|---|---|
| AWS | 0,745 | 2 |
| AWS com Vocabulário Customizado | 0,745 | 3 |
| GCP | 0,624 | 6 |
| GCP com *boost* | 0,638 | 5 |
| Digitro | 0,688 | 4 |
| IBM Watson | 0,801 | 1 |
| MACS | 0,533 | 7 |

Fonte: Elaboração própria (2021).

**Tabela 7 – Comparação WER entre as soluções – Arquivos MP3**

| Soluções | WER Médio (MP3) | Nota |
|---|---|---|
| AWS | 0,756 | 3 |
| AWS com Vocabulário Customizado | 0,756 | 4 |
| GCP | 0,643 | 7 |
| GCP com *boost* | 0,653 | 6 |
| Digitro | 0,684 | 5 |
| IBM Watson | 0,801 | 2 |
| MACS | Não suportado na versão gratuita | 1 |

Fonte: Elaboração própria (2021).

É possível observar que para os arquivos MP3, GCP e GCP com *boost* apresentaram menor WER, seguido pela Digitro, enquanto IBM teve o pior resultado. A solução MACS não converte áudios em MP3 na versão gratuita. Para o formato WAV, MACS e GCP se destacaram e IBM apresentou novamente pior desempenho. Outra observação interessante é que o erro médio é menor para os áudios em WAV.

Vale ressaltar que a adição do vocabulário customizado praticamente não alterou o erro, tanto para AWS, quanto para GCP. Isso ocorre, pois o vocabulário personalizado tem foco nas palavras-chave e, apesar de ajudar na identificação de termos específicos, esses têm pouca representatividade num texto em que a maioria das palavras são artigos, pronomes e preposições.

Pensando em contornar as limitações do WER, que considera apenas a presença ou ausência das palavras, sem considerar sua relevância para a mensagem, as transcrições foram comparadas à luz do KER, um indicador semelhante, mas que leva em consideração as palavras-chave mais relevantes no contexto do setor elétrico.

Como nos resultados anteriores foi identificado um ganho de qualidade nos áudios de formato WAV e considerando a limitação da ferramenta do MACS, foi assumido que este formato é mais interessante para fins de transcrição, de forma que todos os testes vindouros consideram transcrições de arquivos de áudio no formato WAV. A Tabela 8 exibe os resultados da comparação pela métrica KER.

**Tabela 8 – Comparação KER entre as soluções**

| Soluções | KER Médio | Nota |
|---|---|---|
| AWS | 0,742 | 3 |
| AWS com Vocabulário Customizado | 0,743 | 2 |
| Digitro | 0,678 | 4 |
| GCP | 0,631 | 6 |
| GCP com *boost* | 0,636 | 5 |
| IBM Watson | 0,803 | 1 |
| MACS | 0,533 | 7 |

Fonte: Elaboração própria (2021).

É possível observar a diferença de desempenho entre as soluções nos cenários das Tabelas 6, 7 e 8, e isto é um indicador de que as soluções têm desempenhos diferentes, se comparando todas as palavras transcritas ou apenas aquelas mais relevantes. Apesar disso, a comparação por WER e KER ainda carregam consigo as comparações inerentes às palavras que não tem um significado específico, mas ajudam a construir sentenças (*stop words)* e os erros associados às variações de palavras que tem o mesmo significado (*stemming*).

Objetivando realizar uma comparação livre desses vieses, foram realizadas duas outras comparações utilizando o WER e KER. Na primeira é calculado o WER após aplicação do tratamento para remoção das *stop words*, dirimindo a influência desta categoria de palavras nas análises de qualidade de transcrição. Posteriormente

foi computado o KER, após aplicação do *stemming*, cujo objetivo é considerar igualmente palavras com variações diferentes (gênero, número e classe gramatical), mas que conservam o mesmo significado.

As Tabelas 9 e 10 exibem os resultados da comparação do WER com a remoção de *stop words* e a comparação do KER com a aplicação de *stemming*, respectivamente.

**Tabela 9 – Comparação WER com remoção de *stop words***

| Soluções | WER e remoção de *stop words* | Nota |
|---|---|---|
| AWS | 0,758 | 2 |
| AWS com Vocabulário Customizado | 0,756 | 3 |
| Digitro | 0,665 | 4 |
| GCP | 0,636 | 6 |
| GCP com *boost* | 0,637 | 5 |
| IBM Watson | 0,815 | 1 |
| MACS | 0,507 | 7 |

Fonte: Elaboração própria (2021).

**Tabela 10 – Comparação KER com s*temming***

| Soluções | KER com *stemming* | Nota |
|---|---|---|
| AWS | 0,735 | 2 |
| AWS com Vocabulário Customizado | 0,731 | 3 |
| Digitro | 0,642 | 4 |
| GCP | 0,620 | 5 |
| GCP com *boost* | 0,619 | 6 |
| IBM Watson | 0,784 | 1 |
| MACS | 0,784 | 7 |

Fonte: Elaboração própria (2021).

Comparando os dados das Tabelas 9 e 10 é possível observar que no geral, não há diferença significativa no WER com a remoção de *stop words*, mas há uma queda mais acentuada no KER com a aplicação do *stemming*. O que permite concluir que para fins de comparação entre as diferentes soluções de transcrição, as *stop words*, embora indesejáveis, tem um efeito semelhante a um ruído branco. Diminuindo a qualidade das transcrições como um todo, mas sem influenciar especialmente alguma solução, sendo aplicada de maneira quase uniforme a todas elas. O tratamento de *stemming*, em contrapartida, apresentou uma melhora no desempenho das transcrições e seu tratamento fica recomendado para otimizar os resultados alcançados.

A comparação das transcrições utilizando KER, embora ofereça resultados mais contundentes do que a comparação pelo WER, ainda apresenta algumas

limitações. Por exemplo, algumas transcrições, embora apresentem valores insatisfatórios de KER, foram capazes de transmitir parcelas relevantes de informação. Para estes casos, foi evidenciado que a análise pura da quantidade de acertos de palavras, sejam genéricas ou específicas do contexto do setor elétrico, pode não ser o melhor, ou pelo menos único, indicador da qualidade das transcrições.

Pensando em avaliar o valor das informações transcritas, independente das palavras, foi utilizada a técnica de avaliação por reconhecimento de entidades nomeadas *Named Entity Recognition* (NER) (TRAN; BUI, 2021). A técnica consiste em identificar nos textos de hipótese e referência, entidades que possam ser comparadas em tipo, número e valor.

A menção de nomes próprios, tais como nomes de ativos e localidades, é fundamental para a finalidade do uso de modelos StT, pois auxilia na correlação de áudios com os eventos. Dessa forma, identificar quantas entidades presentes no GT foram encontradas nas transcrições dos serviços testados é fator decisivo na validação da qualidade das transcrições.

É importante ressaltar que o treinamento de modelos capazes de identificar corretamente entidades é um tema bastante complexo, e pode ser o foco de outros trabalhos cujo objetivo seja rotular ou classificar ligações. Para os fins de avaliação da qualidade das soluções de transcrição, foi utilizado um modelo minimamente treinado. Embora sua eficiência total seja baixa, como o mesmo modelo é utilizado para diferentes ferramentas, a comparação entre os resultados é preservada e o objetivo de comparação do desempenho das transcrições é atingido.

Para a comparação das soluções foram utilizados dois tipos de entidades: localização (LOC), representando nomes de cidades e bairros; e organização (ORG), representando nomes de usinas e subestações da ENGIE. Por fim, foi realizada uma comparação geral, considerando a quantidade total de entidades, independentemente do tipo. As Tabelas 11, 12 e 13 exibem os resultados de comparação de NER por localização, organização e total, respectivamente.

**Tabela 11 – Comparação NER entre localidades**

| Soluções | Entidades LOC [%] | Nota |
|---|---|---|
| AWS | 0,373 | 1 |
| AWS com Vocabulário Customizado | 0,757 | 7 |
| Digitro | 0,553 | 3 |
| GCP | 0,662 | 5 |
| GCP com *boost* | 0,688 | 6 |
| IBM Watson | 0,474 | 2 |
| MACS | 0,557 | 4 |

Fonte: Elaboração própria (2021).

**Tabela 12 – Comparação NER entre organizações**

| Soluções | Entidades ORG [%] | Nota |
|---|---|---|
| AWS | 0,056 | 1 |
| AWS com Vocabulário Customizado | 0,352 | 7 |
| Digitro | 0,306 | 6 |
| GCP | 0,25 | 4 |
| GCP com *boost* | 0,222 | 3 |
| IBM Watson | 0,12 | 2 |
| MACS | 0,259 | 5 |

Fonte: Elaboração própria (2021).

**Tabela 13 – Comparação NER entre todas as entidades**

| Soluções | Entidades Totais [%] | Nota |
|---|---|---|
| AWS | 0,431 | 3 |
| AWS com Vocabulário Customizado | 0,042 | 6 |
| Digitro | 0,463 | 2 |
| GCP | 0,046 | 5 |
| GCP com *boost* | 0,005 | 7 |
| IBM Watson | 0,472 | 1 |
| MACS | 0,296 | 4 |

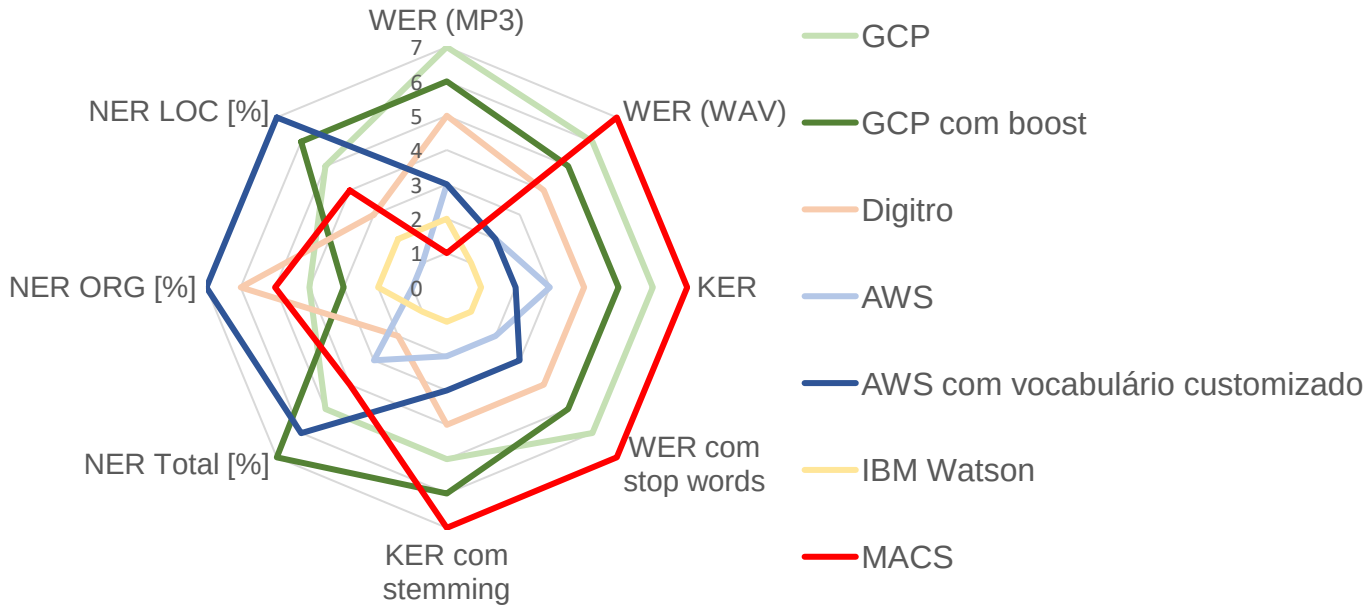
Fonte: Elaboração própria (2021).

Ao avaliar as soluções sob estes critérios, foi possível confirmar o potencial de melhoria da qualidade das transcrições por meio da personalização dos vocabulários, ou configuração de *boost*. Para as identificações das entidades LOC e ORG, a solução da AWS padrão teve o pior desempenho, enquanto a AWS com vocabulário personalizado teve o melhor reconhecimento de entidades dentre todas as soluções. A solução do GCP também teve resultados promissores, mas quando

aplicados os *boosts*, desempenhou muito bem na identificação das entidades LOC e total. A solução MACS, que apresentou menor WER e KER, ficou aquém de ambos os serviços ao utilizarmos o NER. Desta forma, a análise de entidades revelou padrões diferentes que WER e KER não são capazes de identificar, além de comprovar o potencial de otimização das transcrições, utilizando as funcionalidades de personalização das soluções.

A fim de visualizar a comparação entre as sete soluções avaliadas, de acordo com os oito critérios apresentados, a Figura 11 exibe o gráfico de radar das notas de cada solução, para cada critério. O objetivo é fornecer um suporte visual para comparação das soluções, em que a melhor ferramenta seria aquela com maior raio no radar.

**Figura 11 - Comparação entre soluções de transcrição**




Fonte: Elaboração própria (2021).

Observando a Figura 11 é possível identificar que nenhuma solução foi completamente dominada ou dominou todas as outras. Apesar disto, a solução da IBM Watson não apresentou um bom desempenho em nenhum dos critérios avaliados. A solução da Dígitro, embora desenvolvida nativamente para o idioma português, apresentou alguns resultados intermediários, embora nenhum excepcional. Esses resultados corroboram para que ambas as soluções não sejam consideradas as mais promissoras, para este contexto específico.

A solução do MACS foi a mais bem colocada no que tange à identificação de palavras, seja WER ou KER, além de apresentar resultados sólidos para o NER. A observação importante é sobre a limitação do uso de arquivos MP3 para as versões gratuitas da solução. Nos resultados dos testes das Tabelas 6 e 7, entretanto, foi identificada uma melhora no desempenho das transcrições provenientes de arquivos WAV, de sorte que para soluções que armazenem áudios neste formato, ou possam dispor de etapa específica de conversão, a limitação do MACS não seja um impedimento.

As soluções do GCP foram as mais equilibradas, apresentando um bom desempenho em todos os critérios. A utilização do *boost*, entretanto, não melhorou substancialmente a qualidade das transcrições, sendo que em alguns casos, a versão padrão da solução apresentou resultados melhores que após a personalização. Este comportamento atesta para a qualidade da solução, mas é insuficiente no tocante à possibilidade de exploração por usuários mais avançados, dispostos a investir mais tempo para otimizar os modelos de transcrição.

A solução da AWS, em contrapartida, apresentou nativamente um dos resultados menos interessantes dentre todas as opções. Após a configuração do vocabulário personalizado, entretanto, o desempenho de reconhecimento de entidades foi o melhor dentre todas as soluções. Isto atesta que embora a solução não seja nativamente tão adaptada ao português e, sobretudo ao setor elétrico, suas possibilidades de personalização são muitas, permitindo que usuários avançados possam usufruir de bons resultados, desde que dispostos a investir na configuração de vocabulários especializados.

Como conclusão, baseado nos estudos aqui reportados, as soluções MACS, GCP e AWS são as mais indicadas hoje para se trabalhar com transcrições de áudios dos centros de controle do setor elétrico. Embora tenham características próprias, essas soluções apresentaram resultados sólidos e podem ser empregadas com níveis satisfatórios de qualidade das transcrições. Por fim, um aspecto relevante a considerar é sobre os serviços de armazenamento dos arquivos. Cada uma das soluções usa sua própria infraestrutura em nuvem para armazenar os arquivos de áudio antes e após a transcrição, de sorte que eventuais sinergias ou limitações devam ser consideradas, durante a tomada de decisão.

Após os testes e análises apresentados, para os fins desta dissertação, foi decidido usar a solução da AWS, por dois principais motivos:

a) alto desempenho de transcrições com dicionário personalizado. Um dos objetivos deste trabalho é desenvolver dicionários personalizados para o setor elétrico, que possam aumentar a qualidade das transcrições dos áudios dos centros de operação. Baseado na comparação entre o desempenho das transcrições da AWS nativamente, e após a utilização de um dicionário simplificado, foi possível comprovar a flexibilidade da ferramenta, e serão exploradas as possibilidades de criação de novos vocabulários;
b) sinergia com a infraestrutura da ENGIE. Embora a dissertação tenha sido desenvolvida com recursos próprios, usando contas pessoais e limites gratuitos, o projeto de P&D ANEEL é desenvolvido utilizando a infraestrutura de nuvem da ENGIE, que utiliza os serviços da AWS. Um dos objetivos é aproveitar o conhecimento criado neste trabalho para melhorar a qualidade do projeto ANEEL, e utilizar a mesma solução garante o reaproveitamento dos conhecimentos criados.

## 2.7 Criação dos vocabulários personalizados

Esta seção explora as possibilidades de personalização dos vocabulários. Seu intuito é agregar valor para uma ferramenta de StT adaptada para a operação do setor elétrico, que possa transcrever corretamente jargões técnicos, nomes próprios e considerar variações de pronuncias. Para estes fins, é possível fornecer à AWS informações sobre como processar os arquivos de áudio de entrada, criando vocabulários personalizados, que são relações de palavras e frases específicas as quais se deseja que a AWS reconheça.

A AWS permite especificar os vocabulários personalizados em formato de lista simples de palavras, ou por tabela de quatro colunas, que oferece mais controle sobre a entrada e a saída das palavras do vocabulário personalizado. Para explorar mais profundamente as opções de personalização dos vocabulários, nesta dissertação será explorada a opção de vocabulário por tabelas (AMAZON, 2021).

Neste formato, cada linha na tabela é uma palavra ou frase e é formada por quatro campos:

a) *Phrase*. É o único campo obrigatório da tabela e representa a palavra ou frase que deve ser reconhecida. Caso a entrada seja uma frase, é necessário separar as palavras com um hífen. “Cana Brava” deve ser representada como “Cana-Brava”, por exemplo. Para acrônimos ou outras palavras cujas letras devem ser pronunciadas individualmente, é necessário representar as siglas com letras únicas seguidas por pontos, como “C.O.G.” ou “O.N.S.”. Para especificar a forma plural de um acrônimo, é possível separar o “s” do acrônimo com um hífen, como “P.G.I.-s”.
b) *DisplayAs*. Campo opcional que define como a palavra ou frase será exibida no formato de texto, após a transcrição. É possível mapear uma *Phrase* como “Usina Hidrelétrica de Itá” para ser exibida apenas como UHIT, por exemplo. Este campo pode ser utilizado para padronizar a representação, sobretudo de termos que apresentam alguma variação, para que as buscas por palavras-chave, classificação de chamadas ou qualquer outro tipo de rastreabilidade possam ser facilmente implementados nas transcrições. Caso a propriedade não seja especificada, a AWS utiliza o valor do campo *Phrase*.
c) AFI. O Alfabeto Fonético Internacional (AFI) é um campo opcional e pode ser utilizado para especificar a pronúncia da palavra ou frase. Embora bastante útil para outros idiomas mais consolidados, esta funcionalidade não está completa para o português, sendo que alguns fonemas do idioma ainda não tem representação, motivo pelo qual este campo não será utilizado nos vocabulários personalizados deste trabalho. Felizmente, a AWS dispões de outro recurso para lidar com os fonemas das palavras e frases.
d) *SoundsLike*. Campo opcional, com o mesmo objetivo do AFI. Para descrever um fonema, entretanto, o *SoundsLike* permite fissionar uma palavra ou frase em partes menores e fornecer uma pronúncia para cada fração, usando a ortografia padrão do idioma para imitar o som da palavra. A pronúncia da palavra “ANEEL”, por exemplo, pode ser descrita como “á-nê-él”. Por serem opções concorrentes, caso especificado o campo *SoundsLike*, não é possível utilizar o campo AFI e vice-versa. Nos vocabulários aqui desenvolvidos, será dada preferência pelo campo *SoundsLike*. Além de poder ser usado para melhorar a identificação de áudios e transcrição das palavras, o *SoundsLike* é fundamental para criar variações de pronúncias e sotaques. Para tanto, basta criar mais de uma linha,

com a mesma *Phrase*, mas com o campo *SoundsLike* distintos, representando diferentes pronúncias (AMAZON, 2021). A Tabela 14 exibe um exemplo de modelagem de diferentes pronúncias para uma mesma palavra. Mecanismo fundamental para otimização das transcrições no setor elétrico com dimensões continentais.

**Tabela 14 - Modelando diferentes pronúncias no vocabulário personalizado**

| *Phrase* | *SoundsLike* | *DisplayAs* |
|---|---|---|
| watts | ú-á-tis | watts |
| watts | vá-tis | watts |

Fonte: Elaboração própria (2021).

### 2.7.1 Vocabulários Parciais

Uma vez conhecidas as regras para criação dos vocabulários, foi criada a tabela de vocabulário personalizado. Utilizando a experiência empírica do autor, exercitando a escuta de áudios e transcrevendo manualmente as chamadas, foi possível identificar as palavras para criação do vocabulário. Em sua primeira versão, o vocabulário já contava com mais de 300 termos entre palavras e frases específicas, excetuando-se as variações de pronúncia.

Ao contrário do que pode indicar o senso comum, vocabulários muitos amplos não representam necessariamente um incremento direto na qualidade das transcrições. Na prática, o excesso de termos específicos acaba confundindo o modelo, que usa palavras semelhantes durante a transcrição. Foram encontrados casos em que a versão padrão da AWS transcreveu áudios com qualidade superior ao modelo com um vocabulário super populado.

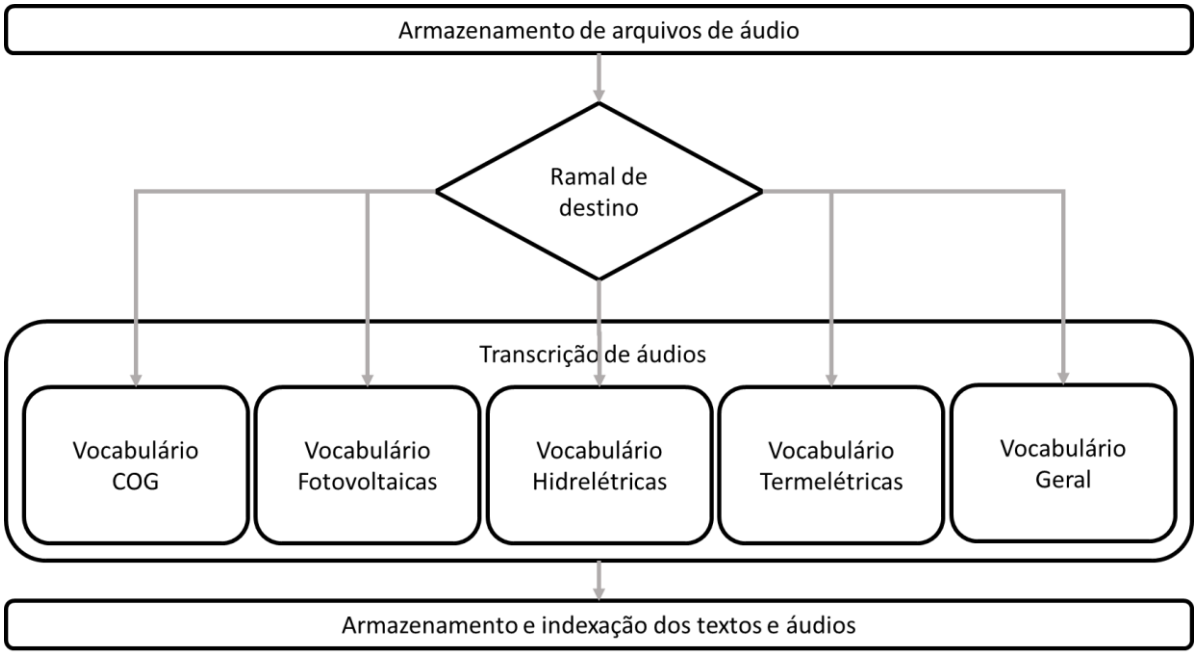
Embora os 300 termos sejam pertinentes para o contexto do setor elétrico, muitos não se aplicavam à maioria das ligações. Termos como “vertedouro”, “vazão” e “jusante”, por exemplo, apenas são relevantes no contexto de chamadas envolvendo hidrelétricas. Em casos de ligações para usinas eólicas ou térmicas, estes jargões acabam apenas poluindo o vocabulário, diluindo a relevância de outras palavras pertinentes e aumentando a incidência de erros.

Objetivando manter a expertise acumulada em um vocabulário extenso, mas sem causar erros por excesso de informações, foi decidido particionar o

vocabulário personalizado em versões menores e mais especializadas. Como critério, foi utilizada a natureza dos ativos de geração que recebem as ligações. Desta forma, foi possível criar um vocabulário personalizado para usinas hidrelétricas, termelétricas, fotovoltaicas, centros de operação e geral. Para definir o vocabulário utilizado para a transcrição, foram consultados os metadados de ramal de destino de cada áudio. A solução se mostrou bastante eficiente, pois com uma lógica condicional simples, foi possível eleger opções de vocabulários personalizados para cada ligação, sem o risco de adicionar complexidade ao processo.

A Figura 12 exibe o fluxo de transcrição dos áudios, utilizando cinco vocabulários personalizados.

**Figura 12 - Transcrição utilizando vocabulários personalizados parciais**




Fonte: Elaboração própria (2021).

A sequência desta seção exibe todos os vocabulários em formato de tabela. É interessante observar que embora a versão final do modelo seja composta por cinco vocabulários personalizados, todos eles compartilham um cerne composto de alguns termos comuns a todos os vocabulários. Estes termos estão presentes em todos os contextos (e vocabulários) e, para evitar repetições, serão apresentados uma única vez. A Tabela 15 exibe os termos comuns a todos os vocabulários, embora ela mesma não seja considerada um vocabulário para fins da solução empregada neste trabalho.

A Tabela 16 exibe o vocabulário personalizado utilizado nas chamadas destinadas ao Centro de Operação da Geração (COG) da ENGIE. Além dos termos comuns a todos os vocabulários, exibidos na Tabela 15, a Tabela 16 possui os nomes próprios dos ativos operados pelo centro de controle, bem como alguns jargões da atividade de operação de usinas hidrelétricas e parques eólicos, ativos predominantes nas chamadas do centro de controle.

**Tabela 15 - Termos comuns aos vocabulários personalizados**

| *Phrase* | *SoundsLike* | *DisplayAs* |
|---|---|---|
| ANEEL | á-nê-él | ANEEL |
| C.O.S. | cê-ó-ésse | COS |
| Engie | en-gi-ê | Engie |
| P.G.I. | pê-gê-í | PGI |
| S.G.I. | ésse-gê-í | SGI |
| ampère | am-pér | ampère |
| ampères | am-pé-res | ampères |
| converter | | converter |
| Diesel | | diesel |
| disjuntor | | disjuntor |
| disponibilizada | | disponibilizada |
| disponível | | disponível |
| emergência | | emergência |
| energizada | | energizada |
| energizado | | energizado |
| energizar | | energizar |
| gerador | | gerador |
| implementa | | implementa |
| implementar | | implementar |
| indisponível | | indisponível |
| intempestivo | | intempestivo |
| Isolar | | isolar |
| isolação | | isolação |
| k.V.Ar | cá-vê-a-é-re | kVAr |
| liberada | | liberada |
| liberado | | liberado |
| Liberar | | liberar |
| limitada | | limitada |
| limitado | | limitado |
| limitar | | limitar |
| manobrar | | manobrar |
| normalizar | | normalizar |

| *Phrase* | *SoundsLike* | *DisplayAs* |
|---|---|---|
| parado | | parado |
| pararráio | | pararráio |
| perturbação | | perturbação |
| programada | | programada |
| proteção | | proteção |
| quilo-volts | qui-lo-vou-tis | quilo-volts |
| reduzir | | reduzir |
| regulador | | regulador |
| restrição | | restrição |
| seccionadora | | seccionadora |
| serviços-auxiliares | | serviços auxiliares |
| sincronismo | | sincronismo |
| sincronizado | | sincronizado |
| sistema | | sistema |
| sistêmica | | sistêmica |
| sobrecorrente | | sobrecorrente |
| sobrefrequência | | sobrefrequência |
| sobretemperatura | | sobretemperatura |
| sobretensão | | sobretensão |
| sobrevelocidade | | sobrevelocidade |
| subcorrente | | subcorrente |
| subestação | | subestação |
| subfrequência | | subfrequência |
| subtensão | | subtensão |
| tensão | | tensão |
| transformador | | transformador |
| transformação | | transformação |
| urgente | | urgente |
| usina | | usina |
| velocidade | | velocidade |
| volts | vou-tis | volts |
| watts | ú-á-tis | watts |

Fonte: Elaboração própria (2021).

**Tabela 16 - Vocabulário personalizado do COG**

| *Phrase* | *SoundsLike* | *DisplayAs* |
|---|---|---|
| Assú | a-sú | Assú |
| black-start | | black start |
| Cacimbas | ca-cim-bas | Cacimbas |
| Campo-Largo | cam-po-lar-go | Campo Largo |
| Cana-Brava | ca-na-bra-va | Cana-Brava |
| C.O.G. | cê-ó-gê | COG |
| compensador | | Compensador |
| compensar | | Compensar |
| conjunto-eólico | com-jun-to-ê-ó-li-co | conjunto eólico |
| depleciona | | Depleciona |
| deplecionar | | Deplecionar |
| Flecheiras | fle-xê-ras | Flecheiras |
| Grid | gri-di | Grid |
| Guajirú | gua-ji-rú | Guajirú |
| hidrelétrica | | hidrelétrica |
| Jaguara | já-gua-ra | Jaguara |
| José-Gelázio | jo-zé-gê-lá-zio | José Gelázio |
| jusante | | jusante |
| Machadinho | ma-cha-di-nho | Machadinho |
| Miranda | mi-ram-da | Miranda |
| montante | | montante |
| Mundaú | mun-da-ú | Mundaú |
| nível-do-reservatório | | nível-do-reservatório |
| Ouro-Verde | ou-ro-ver-de | Ouro Verde |
| parque-eólico | par-kê-ê-ó-li-co | parque eólico |
| Passo-Fundo | pa-so-fun-do | Passo Fundo |
| P.C.H. | pê-cê-ágá | PCH |
| Ponte-de-Pedra | pon-te-de-pe-dra | Ponte de Pedra |
| reservatório | | reservatório |
| Rondonópolis | ron-do-nó-po-lis | Rondonópolis |
| Salto-Santiago | sal-to-san-ti-a-go | Salto Santiago |
| Santa-Mônica | san-ta-mô-ni-cá | Santa Mônica |
| São-Salvador | são-sal-va-dor | São Salvador |
| síncrono | | síncrono |
| sobreposição | so-bre-po-si-ção | sobreposição |
| Trairi | tra-í-ri | Trairi |
| turbina | | turbina |
| Umburanas | um-bu-râ-nas | Umburanas |
| vertimento | | vertimento |

Fonte: Elaboração própria (2021).

A Tabela 17 exibe os termos do vocabulário personalizado utilizado nas chamadas destinadas às usinas hidrelétricas. Além de nomes próprios e procedimentos, são elencados componentes deste tipo de ativo.

**Tabela 17 - Vocabulário personalizado de usinas hidrelétricas**

| ***Phrase*** | ***SoundsLike*** | ***DisplayAs*** |
|---|---|---|
| auto-reestabelecimento | áu-to-res-ta-be-le-ci-men-to | auto-reestabelecimento |
| black-start | | black start |
| compensador | | compensador |
| compensar | | compensar |
| depleciona | | depleciona |
| deplecionar | | deplecionar |
| Estreito | es-trei-to | Estreito |
| hidrelétrica | | hidrelétrica |
| Itá | i-tá | Itá |
| jusante | | jusante |
| montante | | montante |
| nível-do-reservatório | | nível-do-reservatório |
| P.C.H. | pê-cê-ágá | PCH |
| reduz-o-vertimento | | reduz o vertimento |
| reservatório | | reservatório |
| Salto-Osório | sal-to-o-só-rio | Salto Osório |
| síncrono | | síncrono |
| sobreposição | so-bre-po-si-ção | sobreposição |
| turbina | | turbina |
| vertido | | vertido |
| vertimento | | vertimento |

Fonte: Elaboração própria (2021).

De maneira análoga, a Tabela 18 exibe os termos do vocabulário personalizado utilizado nas chamadas destinadas às usinas termelétricas. Além de nomes próprios dos ativos termelétricos, são elencadas alguns componentes e partes deste tipo de usina.

**Tabela 18 - Vocabulário personalizado de usinas termelétricas**

| *Phrase* | *SoundsLike* | *DisplayAs* |
|---|---|---|
| biomassa | | biomassa |
| caldeira | | caldeira |
| Co-geração | co-ge-ra-ção | Co-geração |
| Complexo-termelétrico | com-ple-xo-ter-me-lé-tri-co | Complexo Termelétrico |
| emissão-de-particulado | | emissão de particulado |
| Ferrari | | Ferrari |
| Ibitiúva | i-bi-ti-ú-va | Ibitiúva |
| Jorge-Lacerda | jor-ge-la-cer-da | Jorge Lacerda |
| Lacerda | la-cer-da | Lacerda |
| Lages | la-ges | Lages |
| limpeza-do-condensador | | limpeza do condensador |
| Pampa | pam-pa | Pampa |
| pampa-sul | pam-pa-sul | Pampa Sul |
| precipitador | | precipitador |
| sobreposição | so-bre-po-si-ção | sobreposição |
| termelétrica | | termelétrica |
| turbina | | turbina |
| U.T.L.A. | ú-tê-éle-á | UTLA |
| U.T.L.B. | ú-tê-éle-bê | UTLB |
| U.T.L.C. | ú-tê-éle-cê | UTLC |

Fonte: Elaboração própria (2021).

A Tabela 19 exibe o vocabulário personalizado utilizado pelas chamadas às unidades fotovoltaicas do grupo. É interessante observar que um tipo de geração ainda em estágios de maturidade, seja representado por um vocabulário menor, considerando poucos nomes de ativos e apenas alguns termos técnicos específicos.

**Tabela 19 - Vocabulário personalizado de usinas fotovoltaicas**

| *Phrase* | *SoundsLike* | *DisplayAs* |
|---|---|---|
| A.E.G. | á-é-gê | AEG |
| Aurora | au-ro-ra | Aurora |
| conjunto-eólico | com-jun-to-ê-ó-li-co | conjunto eólico |
| Eólico | | eólico |
| Fotovoltaica | | |
| Grid | gri-di | Grid |
| Nova-Aurora | no-va-au-ro-ra | Nova Aurora |
| painél | | painél |
| Tubarão | tu-ba-rão | Tubarão |

Fonte: Elaboração própria (2021).

Por fim, a Tabela 20 exibe o vocabulário personalizado geral, utilizado para chamadas genéricas, não contemplando ramais ou condições previstos nos casos anteriores. De todas as opções, é o maior vocabulário e por razões de excesso de informações, deve ser utilizadoncomo última opção.

**Tabela 20 - Vocabulário personalizado das chamas em geral**

| *Phrase* | *SoundsLike* | *DisplayAs* |
|---|---|---|
| Assú | a-sú | Assú |
| Aurora | au-ro-ra | Aurora |
| biomassa | | biomassa |
| black-start | | black start |
| C.O.G. | cê-ó-gê | COG |
| Cacimbas | ca-cim-bas | Cacimbas |
| caldeira | | caldeira |
| Campo-Largo | cam-po-lar-go | Campo Largo |
| Cana-Brava | ca-na-bra-va | Cana-Brava |
| Co-geração | co-ge-ra-ção | Co-geração |
| compensador | | compensador |
| compensar | | compensar |
| Complexo-termelétrico | com-ple-xo-ter-me-lé-tri-co | Complexo Termelétrico |
| conjunto-eólico | com-jun-to-ê-ó-li-co | conjunto eólico |
| depleciona | | depleciona |
| deplecionar | | deplecionar |
| emissão-de-particulado | | emissão de particulado |
| eólico | | eólico |
| Estreito | es-trei-to | Estreito |
| Estrela | es-tre-la | Estrela |
| Ferrari | | Ferrari |
| Flecheiras | fle-xê-ras | Flecheiras |
| fotovoltaica | | fotovoltaica |
| Grid | gri-di | Grid |
| Guajirú | gua-ji-rú | Guajirú |
| hidrelétrica | | hidrelétrica |
| Ibitiúva | i-bi-ti-ú-va | Ibitiúva |
| Itá | i-tá | Itá |
| Jaguara | já-gua-ra | Jaguara |
| Jorge-Lacerda | jor-ge-la-cer-da | Jorge Lacerda |
| José-Gelázio | jo-zé-gê-lá-zio | José Gelázio |
| jusante | | jusante |
| Lacerda | la-cer-da | Lacerda |
| Lages | la-ges | Lages |
| limpeza-do-condensador | | limpeza do condensador |
| Machadinho | ma-cha-di-nho | Machadinho |

| *Phrase* | *SoundsLike* | *DisplayAs* |
|---|---|---|
| Miranda | mi-ram-da | Miranda |
| montante | | montante |
| Mundaú | mun-da-ú | Mundaú |
| nível-do-reservatório | | nível-do-reservatório |
| Nova-Aurora | no-va-au-ro-ra | Nova Aurora |
| O.N.S. | ó-enê-ésse | ONS |
| Ouro-Verde | ou-ro-ver-de | Ouro Verde |
| P.C.H. | pê-cê-ágá | PCH |
| Pampa | pam-pa | Pampa |
| pampa-sul | pam-pa-sul | Pampa Sul |
| parque-eólico | par-kê-ê-ó-li-co | parque eólico |
| Passo-Fundo | pa-so-fun-do | Passo Fundo |
| Ponte-de-Pedra | pon-te-de-pe-dra | Ponte de Pedra |
| precipitador | | precipitador |
| reservatório | | reservatório |
| Rondonópolis | ron-do-nó-po-lis | Rondonópolis |
| Salto-Osório | sal-to-o-só-rio | Salto Osório |
| Salto-Santiago | sal-to-san-ti-a-go | Salto Santiago |
| Santa-Mônica | san-ta-mô-ni-cá | Santa Mônica |
| São-Salvador | são-sal-va-dor | São Salvador |
| síncrono | | síncrono |
| sobreposição | so-bre-po-si-ção | sobreposição |
| termelétrica | | termelétrica |
| Trairi | tra-í-ri | Trairi |
| Tubarão | tu-ba-rão | Tubarão |
| turbina | | turbina |
| U.T.L.A. | ú-tê-éle-á | UTLA |
| U.T.L.B. | ú-tê-éle-bê | UTLB |
| U.T.L.C. | ú-tê-éle-cê | UTLC |
| Umburanas | um-bu-râ-nas | Umburanas |
| vertido | | vertido |
| vertimento | | vertimento |

Fonte: Elaboração própria (2021).

# 3 PROPOSIÇÃO DE METODOLOGIA

Neste capítulo são condensadas as soluções adotadas durante o desenvolvimento do trabalho. No primeiro momento é apresentada a arquitetura proposta, desde a gravação dos arquivos de áudio, as etapas de transcrição e as aplicações que fazem uso dos áudios transcritos. Em seguida é sumarizada a metodologia de avaliação da qualidade das transcrições, que pode ser utilizada para definir modelos de transcrição, bem como comparar sua evolução ao longo do tempo. Por fim são apresentadas algumas sugestões de aplicações que podem ser exploradas em futuros trabalhos.

## 3.1 Arquitetura proposta

A Figura 13 fornece um auxílio visual da arquitetura da metodologia desenvolvida. Na primeira camada são armazenados, por força de legislação, os arquivos de áudio gravados durante a atividade de operação de ativos do setor elétrico. A contribuição do trabalho nesta etapa diz respeito à gravação das chamadas em alta qualidade. Para isto, podem ser configurados parâmetros do gravador, ou mesmo a busca por soluções que permitam armazenar os áudios com uma taxa de amostragem de 16 kHz.

A etapa seguinte lida com o pré-processamento destes arquivos de áudio, buscando reduzir sua relação sinal-ruído. Apesar de elementares, os tratamentos utilizados apresentaram ganhos significativos e, embora não faça parte do escopo do trabalho, o aprofundamento em técnicas de tratamento de áudio pode contribuir para a qualidade final das transcrições. Foram aplicados tratamentos em cinco etapas distintas:

a) Aplicação de ganhos de 2 dB, buscando amplificar e facilitar a compreensão da voz dos interlocutores;
b) Normalização, cujo objetivo é equalizar a voz dos interlocutores;
c) Filtro passa-alta, para remover ruídos de baixa frequência como vibrações mecânicas;
d) Filtro passa-baixa, para remover frequências mais elevadas, que pouco contribuem na fala;

e) Conversão dos arquivos do formato estéreo para mono.

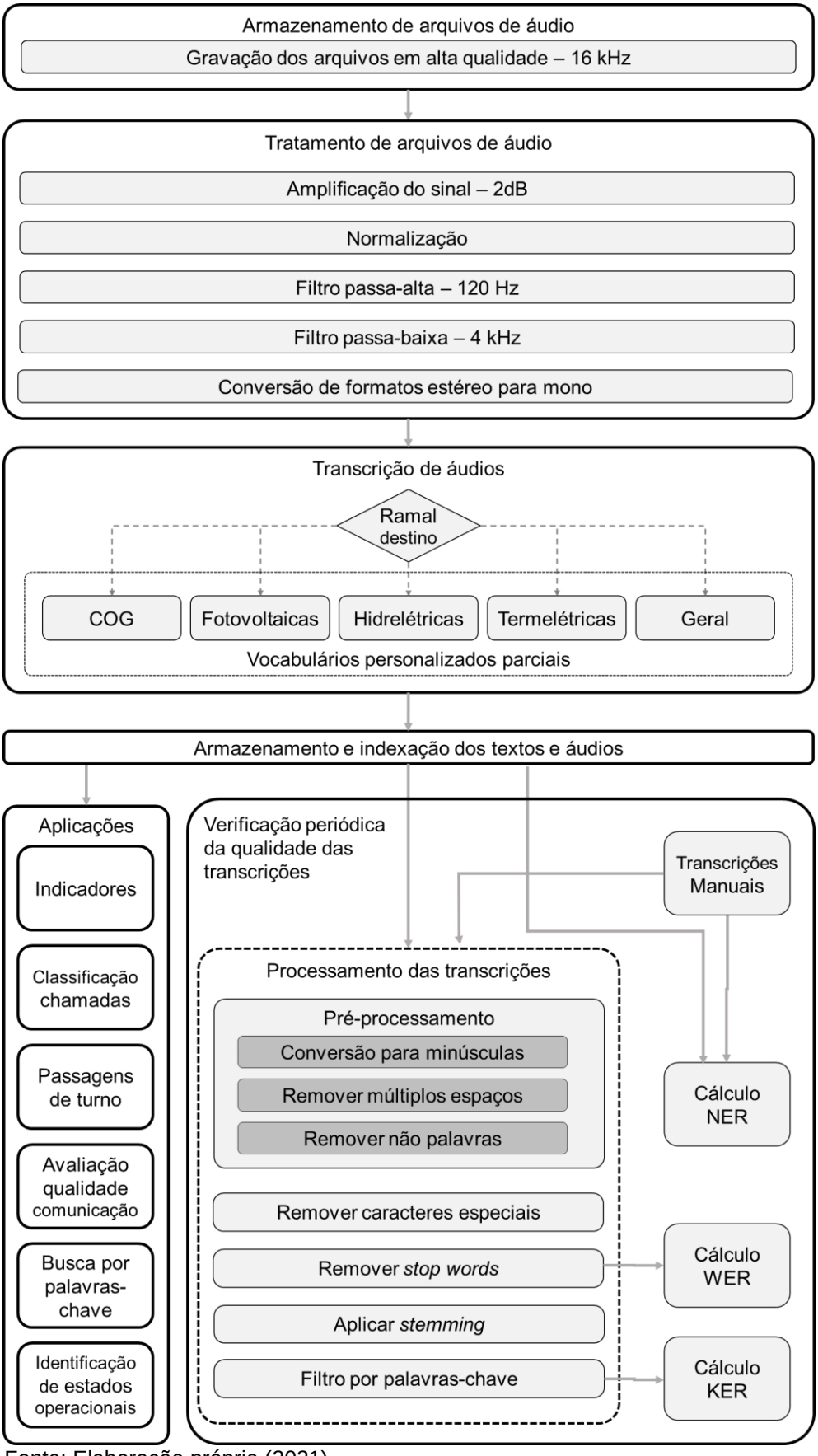
A etapa seguinte é a transcrição de fato e neste ponto, após a pesquisa das principais soluções de mercado, foi eleita a AWS. Ela foi escolhida graças ao desempenho das transcrições e à alta capacidade de personalização dos vocabulários, suportando inclusive variações de pronúncias. Ainda durante os testes de transcrição, foi observado que vocabulários muito extensos tendem a diminuir a qualidade das transcrições e por isso optou-se por uma solução que particiona os vocabulários, de acordo com o tipo de usina. Neste ponto foram criados vocabulários personalizados para o COG, usinas fotovoltaicas, hidrelétricas, termelétricas e geral. O critério para utilização de cada vocabulário são os ramais de destino de cada ligação, armazenados como metadados dos arquivos de áudio.

Após as transcrições os textos são salvos e indexados com seus respectivos áudios. Esta base de dados pode ser utilizada por diferentes aplicações, que utilizam as transcrições indexadas para agregar valor à atividade de operação de ativos do setor elétrico ou para medir o desempenho da solução.

Embora as aplicações não sejam escopo deste trabalho, a Figura 13 mostra como elas se relacionam com o restante da solução. Além disso, foram apresentadas brevemente algumas potenciais aplicações e foi possível identificar que cada uma delas pode ter diferentes relações com as transcrições. Desde aplicações que exigem maior assertividade, até aplicações que desempenhariam bem com uma transcrição não tão precisa. Independentemente das nuances de cada uma, este trabalho se propõe a realizar as transcrições mais fidedignas possíveis, permitindo que cada eventual solução possa escolher a qualidade dos seus insumos.

Por fim, é apresentada a metodologia de avaliação da qualidade das transcrições. Trata-se da mesma metodologia utilizada nesta dissertação para comparação entre as soluções de mercado e para avaliação do impacto das soluções como: tratamento dos arquivos de áudio; desenvolvimento do vocabulário personalizado e criação dos vocabulários parciais. Na Figura 13 o objetivo é utilizar a metodologia para fazer verificações periódicas da qualidade das transcrições, utilizando novas transcrições manuais e monitorar a qualidade do modelo, garantindo que ele não seja deteriorado ao longo do tempo e por eventuais atualizações dos vocabulários personalizados.

**Figura 13 - Arquitetura da metodologia proposta**



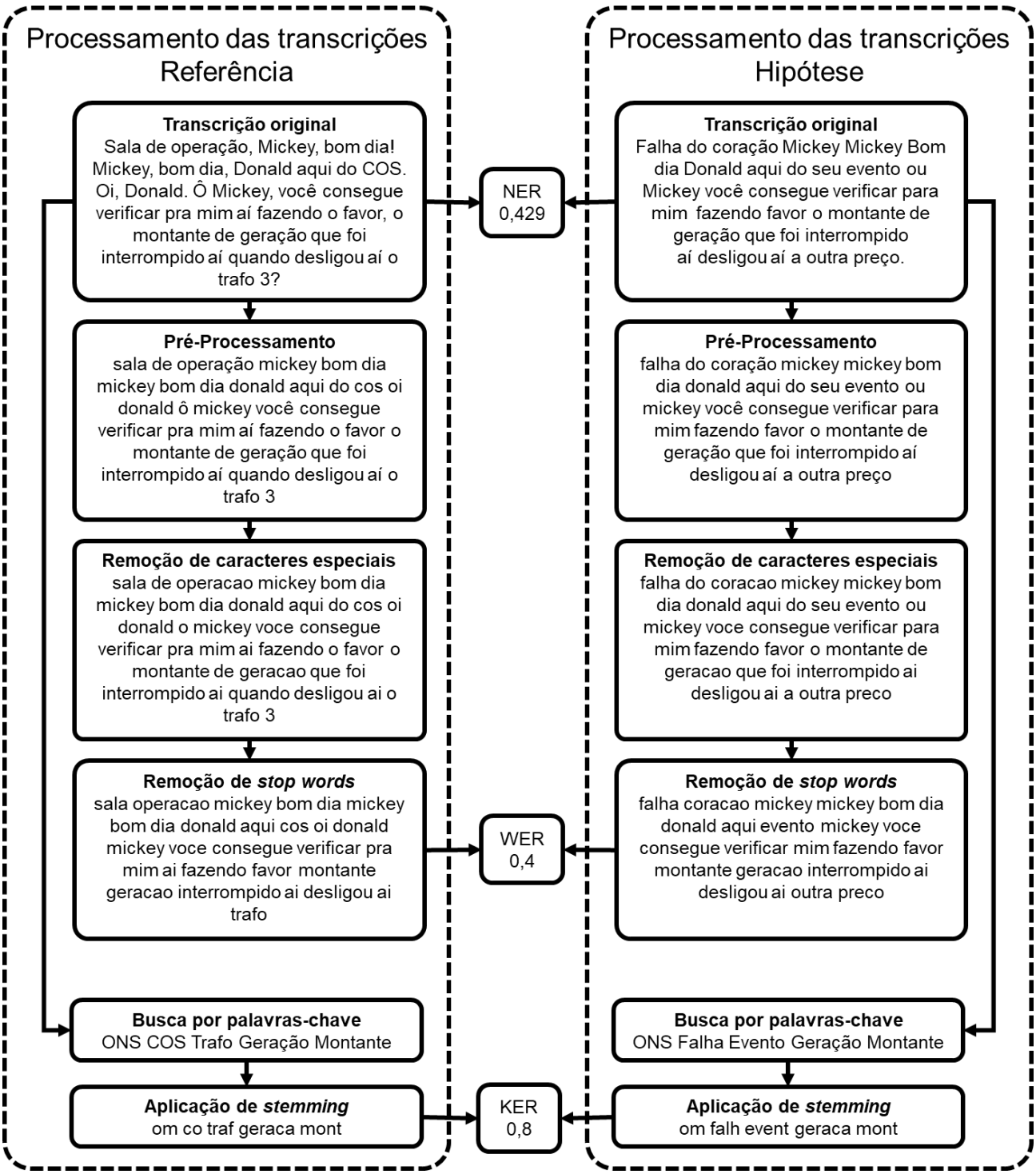

Fonte: Elaboração própria (2021).

## 3.2 Avaliação da qualidade das transcrições

Uma metodologia bastante explorada durante esta dissertação diz respeito à avaliação da qualidade das transcrições. Embora neste trabalho a metodologia tenha sido utilizada para escolha da ferramenta de mercado e definição do vocabulário personalizado, é interessante que a avaliação seja realizada durante toda a vida útil das soluções de transcrição. Este monitoramento permite avaliar o desempenho entre diferentes versões do vocabulário personalizado e a possível degradação da qualidade do modelo.

A Figura 14 exemplifica o tratamento das transcrições para o cálculo do WER, do NER e do KER. Para o cálculo do WER são aplicados os tratamentos básicos, apresentados no APÊNDICE C, que embora possam dificultar a compreensão para leitores humanos, tem por objetivo uniformizar aspectos das transcrições de hipótese e GT antes de compará-las de fato. A métrica NER é baseada na identificação de entidades no texto, de sorte que a comparação é realizada sobre as transcrições originais, com o mínimo de degradação possível. O APÊNDICE H exibe o algoritmo de processamento das transcrições. Por fim, é cálculo do KER leva em consideração as palavras-chave presentes nas transcrições originais, aplicando técnicas de *stemming*. Ao final do processamento a mensagem não tem a estrutura de conversa entre interlocutores humanos, mas preserva a raiz das palavras-chave que dão sentido ao texto original. O APÊNDICE D apresenta os algoritmos de busca por palavras-chave e o APÊNDICE H exibe os algoritmos de processamento das transcrições.

**Figura 14 - Metodologia de tratamento das transcrições - WER, KER e NER**




Fonte: Elaboração própria (2021).

## 3.3 Potenciais aplicações e futuros trabalhos

Como o trabalho foi escrito sob a ótica do setor de geração de energia, discute-se aqui algumas aplicações que podem ser desenvolvidas utilizando como insumo os dados de áudios transcritos. Por terem sido baseados em dados e procedimentos da ENGIE, os itens seguintes refletem algumas das rotinas e desafios enfrentados pela equipe de operação do agente. Esta seção tem por motivação inspirar futuros trabalhos relacionados, apontando tarefas e procedimentos do setor elétrico que podem se beneficiar dos resultados desta dissertação e que podem ser objeto de estudos de iniciativas vindouras.

### 3.3.1 Busca por palavras-chave

Talvez a mais óbvia e útil das aplicações, a busca indexada por palavras-chave tem implementação relativamente simples e um valor imediato. A possibilidade de consultar chamadas de áudio por termos relevantes tem potencial para exponencializar a eficiência das equipes de pós-operação, que são responsáveis, entre outras atividades, por encontrar e enviar evidências, sempre que solicitados, e criação dos RO, RAO (ONS, 2020) e RAP (ONS, 2020). Além disso, outras atividades corriqueiras que necessitam de confirmação via áudio podem se beneficiar. O valor é ainda maior se incrementarmos os filtros com metadados dos arquivos de áudios, como data e duração das chamadas e ramais de origem e de destino.

### 3.3.2 Classificação de chamadas

Embora a quantidade de informações sobre a operação setor elétrico que deva ser reportada via canais de áudio seja grande, a maioria delas é bem padronizada e estruturada (ONS, 2020), de sorte que uma solução capaz de analisar cada uma das transcrições e classificá-las em tipos preexistentes, possivelmente teria uma elevada taxa de assertividade. Os critérios de classificação podem ser distintos. Uma sugestão para futuros trabalhos seria classificar as ligações em rótulos, de acordo com o tema ou temas tratados na ligação: controle de tensão; ajuste de geração; mudança de estado operativo dos geradores; níveis do reservatório; quantidade de chuva; e outros.

### 3.3.3 Avaliação da qualidade da comunicação

O ONS estabelece as regras, os procedimentos básicos e a fraseologia padrão obrigatória que se aplicam às equipes de tempo real dos Centros de Operação do ONS e centros de operação dos agentes, em toda e qualquer comunicação operativa de voz, visando a objetividade, clareza e eficácia das mensagens transmitidas. Os agentes são responsáveis por avaliar, periodicamente, se a qualidade de comunicação dos seus operadores está em conformidade com os critérios elencados no Manual de Procedimento da Operação (MPO) (ONS, 2020). Atualmente, na ENGIE, essa tarefa é realizada de forma manual pela equipe de normatização, que seleciona uma amostra de alguns áudios aleatoriamente de cada operador, usualmente a cada trimestre. Uma sugestão de trabalho futuro é o estudo de viabilidade de uma solução de inteligência computacional que identifique a presença ou ausência de alguns critérios como: identificação do operador; identificação do centro; confirmação da mensagem; clareza e objetividade; assertividade; fraseologia adequada; e cordialidade. Essa solução poderia avaliar uma amostragem muito maior de áudios, automaticamente, evitando o trabalho de seleção e a necessidade de ouvir, atentamente, cada um deles. Além de aumentar a eficiência do processo, esta abordagem permitiria processar uma quantidade maior de chamadas, fornecendo uma estimativa mais precisa da real qualidade da comunicação entre os centros de operação.

### 3.3.4 Passagens de turno

Os centros de operação dos agentes devem dispor de equipes de operação em turnos ininterruptos (ONS, 2020). A operação diária do Centro de Operação do Sistema da ENGIE, por exemplo, é composta por três turnos de 8 horas. Em cada alternância dos operadores acontece a passagem de turno, em que os profissionais que estejam sendo rendidos contextualizem seus substitutos e lhes passem quaisquer informações pertinentes para a operação dos ativos.

Outra aplicação em voga, que faz uso de inteligência computacional, é a sumarização automática de textos. Trata-se de um conceito desenvolvido na década de 1950, que vem despertando interesse da comunidade científica, à medida que as

tecnologias avançam nas áreas de ciências de dados e aprendizado de máquina. Em suma, estas aplicações são capazes de condensar muitas informações, provenientes de várias fontes distintas, em pequenos parágrafos (GAMBHIR; GUPTA, 2016). Uma promissora solução tecnológica seria explorar a possibilidade de desenvolver uma ferramenta que sintetize as ocorrências de cada ativo, mesclando dados transcritos, informações do sistema supervisório (SCADA) e documentações internas em um determinado período, auxiliando a atividade de passagem de turno e destacando condições que demandem mais atenção, contribuindo para a consciência situacional dos operadores.

### 3.3.5 Mudança automática de estados operacionais

Um gerador elétrico conectado no SIN pode transitar entre vários estados operativos. Alguns mais comuns e autoexplicativos são: ligado como gerador (LIG); ligado como compensador síncrono, por solicitação do ONS, para controle de tensão (LCS); ligado como compensador síncrono, por solicitação do agente (LCC); desligado automaticamente por atuação de sistema de proteção ou de controle (DAU); desligado por conveniência operativa do ONS ou por insuficiência de queda útil ou afluência para usinas hidrelétricas (DCO); entre dezenas de outras opções. Estes eventos de mudanças de estado operativo devem ser registrados na base de dados do ONS, por meio de um aplicativo específico, o SAMUG (ONS, 2020).

É responsabilidade do ONS disponibilizar aos agentes de geração, entre outras informações, os dados de mudanças dos estados operativos e disponibilidade de usinas despachadas centralizadamente. Os agentes de geração consistem, sob sua responsabilidade os dados disponibilizados e, por fim, o ONS disponibiliza à Câmara de Comercialização de Energia Elétrica (CCEE) os resultados da apuração dos estados operativos, que serão utilizados para fins de contabilização (ONS, 2020). Quando são encontradas discrepâncias nas apurações do ONS e dos agentes, são consultadas as gravações de áudios dos centros de controle, a fim de confirmar a existência e horários das mudanças dos estados operativos.

Nesse contexto, uma possível área de estudo diz respeito à automação dos registros de mudança de estado operativos dos geradores, através da identificação das transcrições, estruturação das informações e registro automático das chamadas

que tratam deste assunto. Além do impacto financeiro associado e ao ganho de eficiência operacional do processo, soluções deste tipo teriam replicabilidade para a maioria dos agentes conectados no SIN.

### 3.3.6 Indicadores

Esta área de estudo diz respeito a estatísticas e indicadores que podem ser criados ao tratar as transcrições como dados relacionais. É possível encontrar relações não óbvias, cruzando as informações estatísticas das chamadas com outros dados, como escala dos operadores, documentação emitida e dados do sistema supervisório. Questões como sazonalidade das chamadas, quantidade de ligações por tipo, recorrência de comandos, e tipos de solicitações podem ser insumos eficientes para o atendimento regulatório de dimensionamento das equipes dos centros de controle (ONS, 2020).

Uma segunda categoria de indicadores pode ser relacionada com a qualidade da própria ferramenta de transcrição. WER médio por ativos, por exemplo, pode trazer à superfície problemas relacionados à qualidade da telefonia, ruídos ou outras questões de captação. Oscilações na qualidade das transcrições, pode significar algum tipo de degradação do modelo. Incidência alta de chamados com durações longas pode ser um indício de que haja muitas ligações não padronizadas, como conversas com as equipes de manutenção durante intervenções. As possibilidades são inúmeras e a exploração e definição de indicadores é um bom tema a ser explorado em trabalhos futuros.

# 4 CONCLUSÃO

A forma de operar o sistema elétrico tem mudado rapidamente. Tal dinamismo é necessário para manter a confiabilidade da operação, frente aos novos desafios, como por exemplo: a teleassistência de usinas e subestações (ONS, 2020); o planejamento energético com variadas fontes intermitentes (PINHEIRO, 2020); e busca por eficiência e competitividade, em específico quanto à programação do despacho que migra para intervalos semi-horários (FERNANDEZ, 2019).

Apesar dos benefícios, como transparência e reprodutibilidade dos processos e redução da ocorrência de geração fora da ordem de mérito, estas mudanças afetam as equipes de operação de tempo real, que precisam ajustar a geração com maior frequência; e as equipes de pré-operação, que necessitam acordar o programa de geração, junto ao ONS, para o dia seguinte (ONS, 2021).

Neste contexto, é possível reconhecer o potencial da metodologia proposta: transcrições automatizadas de áudios para centros de operação do setor elétrico. Em relação aos insumos, a própria legislação vigente (Procedimentos de Rede) determina que a comunicação via canais de voz seja amplamente utilizada, o que acarreta a disponibilidade de dados relevantes que podem ser transformados em informação útil para auxiliar nas tomadas de decisão para a operação de ativos do setor elétrico.

No que tange à tecnologia, o Capítulo 2 foi capaz de demonstrar elevada assertividade de transcrição, considerando linguagem natural e vocabulário técnico, utilizando ferramentas consolidadas de mercado e explorando ao máximo suas potencialidades. Para isto, foi necessário desenvolver vocabulários especializados, com jargões, termos e nomes próprios do setor elétrico. Os resultados obtidos permitem aplicar inteligência aos textos automaticamente transcritos, conservando a possibilidade para melhorias e evoluções, a depender da retroalimentação dos usuários, garantindo que a qualidade das transcrições seja incremental.

Seja simplificando a execução ou aumentando a qualidade de tarefas já realizadas pelas equipes de operação, ou mesmo proporcionando análises até então inviáveis, as aplicações apresentadas no Capítulo 3 demonstraram grande potencial e podem ser objetos de trabalhos futuros. Utilizando técnicas de inteligência computacional, é possível cruzar os dados transcritos com as informações de outros

sistemas operativos, como sistemas supervisórios e gestão de ocorrências, a fim de gerar relações valiosas para a atividade de operação do setor elétrico.

Para o futuro, podem ser previstas melhorias e evoluções na máquina de transcrição, que pode ser treinada pelos próprios usuários, à medida em que os textos sejam eventualmente corrigidos. É esperado que estes vocabulários possam ser aplicados a outros agentes, ou mesmo outros contextos do setor elétrico.

# REFERÊNCIAS


AFONSO, S. *et al.* Floresta Sintá(c)tica: A treebank for Portuguese, p. 20, abr. 2002.

ALVARES, R. V.; GARCIA, A. C. B.; FERRAZ, I. STEMBR: A Stemming Algorithm for the Brazilian. **Progress in Artificial Intelligence**, v. 3808, p. 693-701, dez. 2005.

AMAZON. **Amazon Transcribe**. Disponivel em: https://aws.amazon.com/transcribe/. Acesso em: 17 maio 2021.

AMAZON. **Custom vocabularies**. Disponivel em: https://docs.aws.amazon.com/transcribe/latest/dg/how-vocabulary.html. Acesso em: 17 maio 2021.

ANEEL. **Geração Distribuída**. Disponivel em: https://app.powerbi.com/view?r=eyJrIjoiZjM4NjM0OWYtN2IwZS00YjViLTllMjItN2E5MzBkN2ZlMzVkIiwidCI6IjQwZDZmOWI4LWVjYTctNDZhMi05MmQ0LWVhNGU5YzAxNzBlMSIsImMiOjR9. Acesso em: 23 mar. 2021.

AUDACITY. **Audacity**. Disponivel em: https://www.audacityteam.org/. Acesso em: 11 abr. 2021.

AUDHKHASI, K. *et al.* Recent progress in deep end-to-end models for spoken language processing. **IBM Journal of Research and Development**, v. 61, p. 1-2, jul. 2017. ISSN 10.1147/JRD.2017.2701207.

AWS. **Improving domain-specific transcription accuracy with custom language models**. Disponivel em: https://docs.aws.amazon.com/transcribe/latest/dg/custom-language-models.html. Acesso em: 17 maio 2021.

BANERJEE, S.; LAVIE, A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements. **43rd Annual Meeting of the Assoc. for Computational Linguistics: Proc. Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization**, Ann Arbor, p. 65-72, jun. 2005.

BATES, M.; WEISCHEDEL, R. M. **Challenges in Natural Language Processing**. [S.l.]: Cambridge University Press, 1993.

BJARGHOV, S. *et al.* Developments and Challenges in Local Electricity Markets: A Comprehensive Review. **IEEE Access**, v. 9, p. 58910-58943, abr. 2021. ISSN 10.1109/ACCESS.2021.3071830.

CHEN, B. *et al.* Networked Microgrids for Grid Resilience, Robustness, and Efficiency: A Review. **IEEE Transactions on Smart Grid**, v. 12, p. 18-32, jan. 2021. ISSN doi: 10.1109/TSG.2020.3010570.

CUISINIER, E. *et al.* Techno-economic planning of local energy systems through optimization models: a survey of current methods. **International Journal of Energy Research**, v. 45, p. 4888-4931, nov. 2020. ISSN doi.org/10.1002/er.6208.

DIAS, M. V. X.; BOROTNI, E. D. C.; JAMIL HADDAD. Geração distribuída no Brasil: oportunidades e barreiras. **Revista Brasileira de Energia**, v. 11, n. Sociedade Brasileira de Planejamento Energético, p. 11, 2015.

DÍGITRO. **Retextoar: transcrição rápida e eficiente de conteúdos**. Disponivel em: https://www.digitro.com/solucao/retextoar-transcricao-rapida-e-eficiente-de-conteudos. Acesso em: 17 maio 2021.

DODDINGTON, G. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. **ARPA Workshop on Human Language Technology**, San Diego, p. 128-132, 2002.

FERNANDEZ, V. D. S. S. Análise do Impacto Financeiro da transição dos modelos de formação de preço DECOMP para o DESSEM nos contratos do ACL: Um estudo de caso com dois perfis de consumo – Industrial e Comercial. **Dissertação (mestrado) - Universidade Federal de Santa Catarina, Departamento de Engenharia Elétrica**, Florianópolis - SC, 2019. 110.

GAMBHIR, M.; GUPTA, V. Recent automatic text summarization techniques. **Artificial Intelligence Review**, v. 47, n. 1, p. 29, mar. 2016.

GOOGLE. **Improve transcription accuracy with speech adaptation boost**. Disponivel em: https://cloud.google.com/speech-to-text/docs/boost. Acesso em: 17 maio 2021.

GOOGLE. **Speech-to-Text: reconhecimento de fala automático**. Disponivel em: https://cloud.google.com/speech-to-text. Acesso em: 17 maio 2021.

IBM. **IBM Watson Speech to Text (STT) - IBM Brasil**. Disponivel em: https://www.ibm.com/cloud/watson-speech-to-text. Acesso em: 17 maio 2021.

IBM. **Idiomas e modelos**. Disponivel em: https://cloud.ibm.com/docs/speech-to-text?topic=speech-to-text-models&locale. Acesso em: 17 maio 2021.

JADON, K. S.; BHADORIA, R. S.; TOMAR, G. S. **International Conference on Computational Intelligence and Communication Networks (CICN)**. doi: 10.1109/CICN.2015.148. [S.l.]: [s.n.]. 2015. p. 727-730.

JIWER. **JiWER: Similarity measures for automatic speech recognition evaluation**. Disponivel em: https://github.com/jitsi/jiwer/. Acesso em: 30 mar. 2021.

KANO, T.; SAKTI, S.; NAKAMURA, S. End-to-End Speech Translation With Transcoding by Multi-Task Learning for Distant Language Pairs. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 28, p. 1342-1355, 2020. ISSN doi: 10.1109/TASLP.2020.2986886.

KLAKOW, D.; PETERS, J. Testing the correlation of word error rate and perplexity. **Speech Communication**, v. 38, p. 19-28, set. 2002.

LEUSCH, G.; UEFFING, N.; NEY, H. DER: Efficient MT Evaluation Using Block Movements. **EACL06**, Trento, p. 241-248, abr. 2006.

MICROSOFT. **Evaluate and improve Custom Speech accuracy**. Disponivel em: https://docs.microsoft.com/pt-br/azure/cognitive-services/speech-service/how-to-custom-speech-evaluate-data. Acesso em: 02 abr. 2021.

MICROSOFT. **Speech to Text**. Disponivel em: https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/. Acesso em: 11 abr. 2021.

MORRIS, A. C.; MAIER, V.; GREEN, P. From WER and RIL to MER and WIL: improved evaluation. **Institute of Phonetics**, 2004.

NADEAU, D.; SEKINE, S. **A survey of named entity recognition and classification**. [S.l.]: Lingvisticæ Investigationes, 2007.

NLTK PROJECT. **NLTK 3.5 documentation**. Disponivel em: https://www.nltk.org/. Acesso em: 03 abr. 2021.

ONS. **Procedimentos de Rede - Submódulo 1.1: O ONS e os procedimentos de rede: visão geral**. Disponivel em: http://www.ons.org.br/%2FProcedimentosDeRede%2FM%C3%B3dulo%201%2FSubm%C3%B3dulo%201.1%2FSubm%C3%B3dulo%201.1_Rev_0.1.pdf. Acesso em: 24 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 1.1: O ONS e os procedimentos de rede: visão geral**. Disponivel em: http://www.ons.org.br/%2FProcedimentosDeRede%2FM%C3%B3dulo%201%2FSubm%C3%B3dulo%201.1%2FSubm%C3%B3dulo%201.1_Rev_0.1.pdf. Acesso em: 24 mar. 2021.

ONS. **Manual de Procedimentos da Operação - Submódulo 5.13 - RO-AO.BR.04: Apuração das Mudanças de Estados Operativos de Unidades Geradoras e Interligações Internacionais**. Disponivel em: http://www.ons.org.br/%2FMPO%2FDocumento%20Normativo%2F4.%20Rotinas%20Operacionais%20-%20SM%205.13%2F4.3.%20Rotinas%20P%C3%B3s-Opera%C3%A7%C3%A3o%2F4.3.2.%20Apura%C3%A7%C3%A3o%20de%20Dados%2FRO-AO.BR.04_Rev.29.pdf. Acesso em: 24 mar. 2021.

ONS. **Manual de Procedimentos da Operação - Submódulo 5.13 - RO-RO.BR.01: Comunicação Verbal na Operação**. Disponivel em: http://www.ons.org.br/%2FMPO%2FDocumento%20Normativo%2F4.%20Rotinas%20Operacionais%20-%20SM%205.13%2F4.1.%20Rotinas%20Gerais%2F4.1.7.%20Relacionamento%20Operacional%2FRO-RO.BR.01_Rev.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Manual de Procedimentos da Operação - Submódulo 5.13 - RO-RO.BR.01: Comunicação Verbal na Operação**. Disponivel em: http://www.ons.org.br/%2FMPO%2FDocumento%20Normativo%2F4.%20Rotinas%20Operacionais%20-%20SM%205.13%2F4.1.%20Rotinas%20Gerais%2F4.1.7.%20Relacionamento%20Operacional%2FRO-RO.BR.01_Rev.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 2.15: Requisitos mínimos para telecomunicações**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%202.15-RQ_2020.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 2.15: Requisitos mínimos para telecomunicações**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%202.15-RQ_2020.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 2.16: Requisitos operacionais para os centros de operação e instalações da Rede de Operação**. Disponivel em: http://www.ons.org.br/%2FProcedimentosDeRede%2FM%C3%B3dulo%2010%2FSubm%C3%B3dulo%2010.14%2FSubm%C3%B3dulo%2010.14%20-%202019.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 2.16: Requisitos operacionais para os centros de operação e instalações da Rede de Operação**. Disponivel em: http://www.ons.org.br/%2FProcedimentosDeRede%2FM%C3%B3dulo%2010%2FSubm%C3%B3dulo%2010.14%2FSubm%C3%B3dulo%2010.14%20-%202019.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 5.1: Operação do Sistema e das instalações da Rede de Operação**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%205.1-RS_2020.12.pdf. Acesso em: 23 mar. 2021.

ONS. P**rocedimentos de Rede - Submódulo 6.2: Análise da operação, ocorrências e perturbações e acompanhamento das providências**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.2-OP_2020.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 6.2: Análise da operação, ocorrências e perturbações e acompanhamento das providências**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.2-OP_2020.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 6.3: Elaboração do Relatório de Análise de Perturbação**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.3-RS_2020.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 6.3: Elaboração do Relatório de Análise de Perturbação**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.3-RS_2020.12.pdf. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 6.5: Apuração da geração e de indisponibilidade de empreendimentos de geração**. Disponivel em:

http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.5-PR_2020.12.pdf. Acesso em: 24 mar. 2021.

ONS. **ENTENDA O QUE MUDA NO ONS COM A ENTRADA EM OPERAÇÃO DO DESSEM EM JANEIRO DE 2020**. Disponivel em: http://www.ons.org.br/Paginas/Noticias/20190807_oquemudacomodessem.aspx. Acesso em: 23 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 7.2: Classificação de modalidade de operação de usinas**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%207.2-PR_2020.12.pdf. Acesso em: 24 mar. 2021.

ONS. **Procedimentos de Rede - Submódulo 7.2: Classificação de modalidade de operação de usinas**. Disponivel em: http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%207.2-PR_2020.12.pdf. Acesso em: 24 mar. 2021.

PAPINENI, K. *et al.* BLEU: a method for automatic evaluation of machine translation. **40th Annual Meeting of the Association for Computational Linguistics (ACL)**, Philadelphia, p. 311-318, jun. 2002.

PARK, Y. *et al*. An Empirical Analysis of Word Error Rate and Keyword Error Rate. **International Conference on Spoken Language Processing**, Brisbane, 2008.

PINHEIRO, V. D. C. N. Despacho Ótimo De Sistemas Elétricos Com Armazenamento De Energia Para Fontes Intermitentes De Geração No Brasil. **Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Civil, Arquitetura e Urbanismo**, Campinas - SP, 2020.

POPOVIC, M.; NEY, H. Word Error Rates: Decomposition over POS Classes and Applications for. **Proceedings of the Second Workshop on Statistical Machine Translation**, Prague, p. 48-55, jun. 2007.

PYSOX. **Welcome to pysox's documentation!**. Disponivel em: https://pysox.readthedocs.io/en/latest/#. Acesso em: 05 abr. 2021.

PYTHON SOFTWARE FOUNDATION. **Python**. Disponivel em: https://www.python.org/. Acesso em: 03 abr. 2021.

PYTHON-LEVENSHTEIN. **Python extension for computing string edit distances and similarities**. Disponivel em: https://pypi.org/project/python-Levenshtein/#documentation. Acesso em: 30 mar. 2021.

RAJARAMAN, A.; ULLMAN, J. D. **Data Mining**. Cambridge: Cambridge University Press, 2011. 1-17 p. ISBN Cambridge University Press.

SCHARENBORG, O. *et al.* Speech Technology for Unwritten Languages. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 28, p. 964-975, 2020. ISSN doi: 10.1109/TASLP.2020.2973896.

SNOVER, M. *et al.* A Study of Translation Error Rate with Targeted Human Annotation. **7th Conf. of the Association for Machine Translation in the Americas (AMTA 06)**, Boston, p. 223-231, 2006.

SPACY. **Industrial-Strength Natural Language Processing**. Disponivel em: https://spacy.io/. Acesso em: 04 abr. 2021.

STROBEL, P. B.; SIMON C, M. V. How Much Data Do You Need? About the Creation of a Ground Truth for Black Letter and the Effectiveness of Neural OCR. **Proceedings of the 12th Conference on Language Resources and Evaluation**, Marseille, p. 3551–3559, nov. 2020.

TORCOLI, M.; KASTNER, T.; HERRE, J. Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 29, p. 1530 - 1541, 2021. ISSN doi: 10.1109/TASLP.2021.3069302.

TRAN, O. T.; BUI, V. T. Neural Text Normalization in Speech-to-Text Systems with Rich Features. **Applied Artificial Intelligence**, v. 35, p. 193-205, jan. 2021. ISSN 10.1080/08839514.2020.1842108.

URGS. RSLP Stemmer (Removedor de Sufixos da Lingua Portuguesa). **Instituto de Informática (INF) - Universidade Federal do Rio Grande do Sul (UFRGS)**. Disponivel em: http://www.inf.ufrgs.br/~viviane/rslp/index.htm. Acesso em: 03 abr. 2021.

USP. Mac-Morpho. **Núcleo Interinstitucional de Lingüística Computacional (NILC), Universidade de São Paulo (USP)**. Disponivel em: http://nilc.icmc.usp.br/macmorpho/. Acesso em: 03 abr. 2021.

WANG, Y.-Y.; ACERO, A.; CHELBA, C. Is word error rate a good indicator for spoken language understanding accuracy. **2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No.03EX721)**, p. 577-582, nov. 2003.

YANG, B. *et al.* Modelling, applications, and evaluations of optimal sizing and placement of distributed generations: A critical state-of-the-art survey. **International Journal of Energy Research**, v. 45, p. 3615-3642, nov. 2020. ISSN doi.org/10.1002/er.6104.

ZECHNER, K.; WAIBEL, A. Minimizing word error rate in textual summaries of spoken language. **North American chapter of the Association for Computational Linguistics**, 2000. 186–193.

# APÊNDICES

# APÊNDICE A – AUTORIZAÇÃO DE DIVULGAÇÃO DE INFORMAÇÕES




**MINISTÉRIO DA EDUCAÇÃO**
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS, DEPARTAMENTO ACADÊMICO DE ELETROTÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

## TERMO DE AUTORIZAÇÃO PARA DIVULGAÇÃO DE INFORMAÇÕES DE EMPRESAS

Empresa: ENGIE BRASIL ENERGIA
CNPJ: 02.474.103/0001-19 Inscrição Estadual: 253.645.700
Representante da empresa: Elinton Chiaradia
*Telefone: (48) 32217368* *E-mail: elinton.chiaradia@engie.com*

Discente: Adrisson Consoni Floriano Matrícula: 201925300003
Título da dissertação: Transcrições automáticas de áudios para centros de operação do setor elétrico
Orientador: Prof. Sérgio Luciano Avila
Coorientador: Prof. Rubipiara Cavalcante Fernandes
Curso/Programa de Pós-graduação: Mestrado Profissional em Sistemas de Energia Elétrica

Tipo de informação e/ou documentos a ser disponibilizados: arquivos de áudios do centro de operação da ENGIE, entre os dias de 01 /01 /2017 e 01 / 05 / 2021.

Como representante da empresa acima nominada, declaro que as informações e/ou documentos disponibilizados pela empresa exclusivamente para o trabalho citado:
( ) podem ser usados e publicados sem restrição.
( ) possuem restrição parcial por um período de ___ anos, não podendo ser publicadas as seguintes informações e/ou documentos: _____
( X ) podem ser utilizados para fins acadêmicos, porém os arquivos não podem ser compartilhados.

Florianópolis, 21 de maio de 2021

DocuSigned by:
Elinton Chiaradia
33871477E1BD4FA...

Representante da Empresa: Elinton Chiaradia

Tendo ciência

DocuSigned by:
66965C54B60242D...

Discente: Adrisson Consoni Floriano

SERGIO LUCIANO AVILA:02074392907
2021.05.21 19:57:45 -03'00'

Orientador: Prof. Sérgio L. Avila

# APÊNDICE B – CÁLCULO DAS MÉTRICAS WER, MER, WIL, WIP

```
import Levenshtein
def calcular_metricas(transcricao_gdt, transcricao_hip):

    transcricao_gdt = remover_caracteres_especiais(
                                    pre_processar_transcricao(transcricao_gdt))
    transcricao_hip = remover_caracteres_especiais(
                                    pre_processar_transcricao(transcricao_hip))

    # Converte todas as palavras em caracteres, pois a biblioteca Levenshtein
    # calculas as substituições, deleções e inserções caractere a caractere e
    # não palavra a palavra
    transc_char_gdt, transc_char_hip = \
                                word_to_char(transcricao_gdt, transcricao_hip)
    operacoes_de_edicao = Levenshtein.editops(transc_char_hip, transc_char_gdt)

    substtdas = sum(1 if op[0] == "replace"else 0 for op in operacoes_de_edicao)
    deletadas = sum(1 if op[0] == "delete" else 0 for op in operacoes_de_edicao)
    inseridas = sum(1 if op[0] == "insert" else 0 for op in operacoes_de_edicao)

    total_palavras_gdt = len(transc_char_gdt)
    total_palavras_hip = len(transc_char_hip)
    corretas = total_palavras_gdt - (substtdas + inseridas)
    edicoes = substtdas + deletadas + inseridas
    wer = edicoes / total_palavras_gdt
    mer = edicoes / (corretas + edicoes)
    wip = (corretas / total_palavras_gdt) * (corretas / total_palavras_hip)
    wil = 1 - wip

    return {
    "Palavras Corretas": corretas,
    "Palavras Totais": total_palavras_gdt,
    "Palavras Deletadas": deletadas,
    "Palavras Inseridas": inseridas,
    "Palavras Substituidas": substtdas,
    "Tamanho Groundtruth": total_palavras_gdt,
    "Tamanho Hipótese": total_palavras_hip,
    "Edições Necessárias": edicoes,
    "% Deleções": deletadas/edicoes,
    "% Inserções": inseridas/edicoes,
    "% Substituições": substtdas/edicoes,
    "WER": wer,
    "MER": mer,
    "WIL": wil,
    "WIP": wip}
```

Função de conversão de palavras para caracteres, chamada no corpo da função principal

```
def word_to_char(transcricao_gdt, transcricao_hip):
    transc_intersecao = set(transcricao_gdt.split() + transcricao_hip.split())
    word2char = dict(zip(transc_intersecao, range(len(transc_intersecao))))
    transc_char_gdt = [chr(word2char[w]) for w in transcricao_gdt.split()]
    transc_char_hip = [chr(word2char[w]) for w in transcricao_hip.split()]
    transc_char_gdt = ''.join(transc_char_gdt)
    transc_char_hip = ''.join(transc_char_hip)
    return transc_char_gdt, transc_char_hip
```

Função de conversão de palavras para caracteres, chamada no corpo da função principal

```
transcricao_gdt = """Sala de operação, Mickey, bom dia! Mickey, bom dia,
                     Donald aqui do COS. Oi, Donald. Ô Mickey, você
                     consegue verificar pra mim aí fazendo o favor, o montante
                     de geração que foi interrompido aí quando desligou aí o
                     trafo 3?"""

transcricao_hip = """Falha do coração Mickey Mickey Bom dia Donald aqui
                     do seu evento ou Mickey você consegue verificar para mim
                     fazendo favor o montante de geração que foi interrompido
                     aí desligou aí a outra preço."""


metricas = calcular_metricas(transcricao_gdt, transcricao_hip)
métricas
```

```
{'% Deleções': 0.0,
 '% Inserções': 0.2727272727272727,
 '% Substituições': 0.7272727272727273,
 'Edições Necessárias': 22,
 'MER': 0.55,
 'Palavras Corretas': 18,
 'Palavras Deletadas': 0,
 'Palavras Inseridas': 6,
 'Palavras Substituidas': 16,
 'Palavras Totais': 40,
 'Tamanho Groundtruth': 40,
 'Tamanho Hipótese': 34,
 'WER': 0.55,
 'WIL': 0.7617647058823529,
 'WIP': 0.23823529411764707}
```

# APÊNDICE C – PRÉ-PROCESSAMENTO DAS TRANSCRIÇÕES

```
import jiwer

def pre_processar_transcricao(transcricao):

    tranformacoes_lista = [
        jiwer.ToLowerCase(),
        jiwer.RemoveMultipleSpaces(),
        jiwer.RemovePunctuation(),
        jiwer.SentencesToListOfWords(),
        jiwer.RemoveEmptyStrings(),
        jiwer.RemoveKaldiNonWords(),
        jiwer.Strip()
        ]

    tratar_transcricao = jiwer.Compose(tranformacoes_lista)
    transcricao_tratada = tratar_transcricao(transcricao)

    return transcricao_tratada
```

## Exemplo de pré-processamento de transcrições

```
transcricao_gdt = """Sala de operação, Mickey, bom dia! Mickey, bom dia,
                     Donald aqui do COS. Oi, Donald. Ô Mickey, você
                     consegue verificar pra mim aí fazendo o favor, o montante
                     de geração que foi interrompido aí quando desligou aí o
                     trafo 3?"""

transcricao_processada = pre_processar_transcricao(transcricao_gdt)
print(" ".join(transcricao_processada))
```

```
sala de operação mickey bom dia mickey bom dia donald aqui do cos oi donald ô mickey
você consegue verificar pra mim aí fazendo o favor o montante de geração que foi
interrompido aí quando desligou aí o trafo 3
```

```
from unidecode import unidecode

def tratar_caracteres_especiais(transcricao):
    transcricao_lista = transcricao.split(" ")

    transcricao_em_unicode = [unidecode(word) for word in transcricao]
    return "".join(transcricao_em_unicode)
```

Exemplo de tratamento de caracteres especiais

```
transcricao_gdt = """Sala de operação, Mickey, bom dia! Mickey, bom dia,
                     Donald aqui do COS. Oi, Donald. Ô Mickey, você
                     consegue verificar pra mim aí fazendo o favor, o montante
                     de geração que foi interrompido aí quando desligou aí o
                     trafo 3?"""

transcricao_em_unicode = tratar_caracteres_especiais(transcricao_gdt)
print(transcricao_em_unicode)
```

```
Sala de operacao, Mickey, bom dia! Mickey, bom dia, Donald aqui do COS. Oi, Donald.
O Mickey, voce consegue verificar pra mim ai fazendo o favor, o montante de geracao
que foi interrompido ai quando desligou ai o trafo 3?
```

# APÊNDICE D – CÁLCULO DA MÉTRICA KER

```
import numpy as np

def calcular_ker(transcricao_gdt, transcricao_hip):

  palavras_chave_hip = buscar_palavras_chave(transcricao_hip)
  palavras_chave_gdt = buscar_palavras_chave(transcricao_gdt)

  acertos_hip = \
            list(set(palavras_chave_gdt).intersection(set(palavras_chave_hip)))
  erros_hip = np.setdiff1d(palavras_chave_hip, palavras_chave_gdt).tolist()
  falta_hip = np.setdiff1d(palavras_chave_gdt, palavras_chave_hip).tolist()
  ker = (len(erros_hip) + len(falta_hip)) / len(palavras_chave_gdt)

  return {
      'Palavras-Chave Hipótese': palavras_chave_hip,
      'Palavras-Chave Groundtruth': palavras_chave_gdt,
      'Hipótese Acertos': acertos_hip,
      'Hipótese Erros': erros_hip,
      'Hipótese Faltas': falta_hip,
      'KER': ker,
      '% Acerto':len(acertos_hip)/len(palavras_chave_gdt) \
                                  if len(palavras_chave_gdt)!=0 else np.nan}

lista_palavras_chave =\
  ["ONS","IHM","ENGIE","PGI","SGI","ANEEL","CNOS","COSR","COS","COG","UHE",
   "PCH","MUST","Sincronismo","Oscilação","Rotação","Turbina","Instabilidade",
   "Manutenção","Alarme","Gerador","Falha","Serviço Auxiliar","Trafo",
   "Partida a Quente","Partida a Frio", "Tomada d'Agua","Fuga Terra","Comando",
   "Evento","Proteção","Transformador","Geração","Programação","Válvula",
   "Partida","Tubulação","Indisponibilidade", "Comporta","Seccionadora",
   "Inversor","Potência Ativa","Potência Reativa", "Fator de Potência","Manobra"
   ,"Retificador","Sobretemperatura","Sobretensão","Sobrecorrente",
   "Sobrefrequência","Sobrevelocidade","Subtensão","Subcorrente","Subfrequência"
   ,"Montante","Jusante","Compensador","Perturbação","Sistêmica","Black start",
   "Para raio","Disjuntor","Energizada","Normalizada","Restrição","Turbina",
   "Caldeira","Precipitador","Vertimento","Vertedouro","Deplecionar","Amperes",
   "Volts","Watts","Hidrelétrica","Parque Eólico","Termelétrica","Aerogerador",
   "Cana Brava","Estreito","Itá","Jaguara","Machadinho","Miranda","Passo Fundo"
   ,"Ponte de Pedra","São Salvador","Salto Osório","Salto Santiago","Lacerda A"
   ,"Lacerda B","Lacerda C","José Gelázio","Rondonópolis","Cacimbas","Estrela",
   "Ouro Verde","Santa Mônica","Flecheiras","Guajirú","Mundaú","Trairi",
   "Campo Largo","Umburanas"]
```

```
def buscar_palavras_chave(transcricao):
  palavras_chave = [i for i in lista_palavras_chave \
                if unidecode(i).lower() in unidecode(transcricao).lower()]
  return palavras_chave
```

Exemplo de busca de palavras-chave e cálculo do KER

```
transcricao_gdt = """Sala de operação, Mickey, bom dia! Mickey, bom dia,
                     Donald aqui do COS. Oi, Donald. Ô Mickey, você
                     consegue verificar pra mim aí fazendo o favor, o montante
                     de geração que foi interrompido aí quando desligou aí o
                     trafo 3?"""

transcricao_hip = """Falha do coração Mickey Mickey Bom dia Donald aqui
                     do seu evento ou Mickey você consegue verificar para mim
                     fazendo favor o montante de geração que foi interrompido
                     aí desligou aí a outra preço."""

palavras_chave = calcular_ker(transcricao_gdt, transcricao_hip)
palavras_chave
```

```
{'% Acerto': 0.6,
 'Hipótese Acertos': ['Montante', 'ONS', 'Geração'],
 'Hipótese Erros': ['Evento', 'Falha'],
 'Hipótese Faltas': ['COS', 'Trafo'],
 'KER': 0.8,
 'Palavras-Chave Groundtruth': ['ONS', 'COS', 'Trafo', 'Geração', 'Montante'],
 'Palavras-Chave Hipótese': ['ONS', 'Falha', 'Evento', 'Geração', 'Montante']}
```

# APÊNDICE E – REMOÇÃO DE *STOP WORDS*

```
import string
import nltk
nltk.download('punkt')
nltk.download('stopwords')

def remover_stopwords(transcricao, min_characters=2):

    transcricao_lista = transcricao.split(" ")

    stopwords = nltk.corpus.stopwords.words("portuguese")
    pontuacao = string.punctuation

    transcricao_sem_stopwords = [palavra for palavra in transcricao_lista
              if (palavra not in stopwords) & (len(palavra) >= min_characters)]


    return " ".join(transcricao_sem_stopwords)
```

## Exemplo de remoção de *stop words*

```
transcricao_gdt = """Sala de operação, Mickey, bom dia! Mickey, bom dia,
                     Donald aqui do COS. Oi, Donald. Ô Mickey, você
                     consegue verificar pra mim aí fazendo o favor, o montante
                     de geração que foi interrompido aí quando desligou aí o
                     trafo 3?"""

transcricao_sem_stopwords = remover_stopwords(transcricao_gdt)
print(transcricao_sem_stopwords)
```

```
Sala operação, Mickey, bom dia! Mickey, bom dia, Donald aqui COS. Oi, Donald.
Mickey, consegue verificar pra mim aí fazendo favor, montante geração interrompido
aí desligou aí trafo 3?
```

# APÊNDICE F – APLICAÇÃO DO *STEMMING*

```
import nltk
nltk.download('rslp')

def aplicar_stemmer(transcricao):
    #RSLPStemmer apresenta erros ao lidar com pontuações, sendo necessário
    #seu tratamento prévio
    transcricao_lista = pre_processar_transcricao(transcricao).split(" ")

    stemmer = nltk.RSLPStemmer()
    transc_com_stemming = [stemmer.stem(word) for word in transcricao_lista]
    return " ".join(transc_com_stemming)
```

## Exemplo de remoção de aplicação de stemming

```
transcricao_gdt = """Sala de operação, Mickey, bom dia! Mickey, bom dia,
                     Donald aqui do COS. Oi, Donald. Ô Mickey, você
                     consegue verificar pra mim aí fazendo o favor, o montante
                     de geração que foi interrompido aí quando desligou aí o
                     trafo 3?"""

transcricao_com_stemming = aplicar_stemmer(transcricao_gdt)
print(transcricao_com_stemming)
```

```
sal de oper mickey bom dia mickey bom dia donald aqu do cos oi donald ô mickey voc
conseg verific pra mim aí faz o favor o mont de ger que foi interromp aí qu deslig
aí o traf 3
```

# APÊNDICE G – CÁLCULO DAS MÉTRICAS NER

```
import pandas
from IPython.core.display import display, HTML
from spacy import displacy

def calcular_metricas_NER(transcricao_gdt, transcricao_hip):
  entidades_hip = buscar_listas_de_entidades(transcricao_hip)
  entidades_gdt = buscar_listas_de_entidades(transcricao_gdt)

  df = criar_df(entidades_gdt, entidades_hip)
  print('_' * 100)
  print("\033[1m" + "Groundtruth")
  html_gdt = displacy.render(entidades_gdt[4], style="ent")
  display(HTML(html_gdt))
  print('_' * 100)

  print("\n"+ "\033[1m" + "Hipótese")
  html_hip = displacy.render(entidades_hip[4], style="ent")
  display(HTML(html_hip))
  print('_' * 100)

  print("\n"+ "\033[1m" + "Quantidade de entridades por tipo")
  df
  return df
```

Função de busca das entidades, chamada no corpo da função principal

```
import spacy
import pt_core_news_sm

def buscar_listas_de_entidades(transcricao):

  nlp = pt_core_news_sm.load()
  doc = nlp(transcricao)
  entidades = doc.ents

  lista_ent_PER = [ent.text for ent in entidades if ent.label_ == 'PER']
  lista_ent_ORG = [ent.text for ent in entidades if ent.label_ == 'ORG']
  lista_ent_LOC = [ent.text for ent in entidades if ent.label_ == 'LOC']
  lista_ent_MIS = [ent.text for ent in entidades if ent.label_ == 'MISC']

  return lista_ent_PER, lista_ent_ORG, lista_ent_LOC, lista_ent_MIS, doc
```

Função criação do resultado em tabela, chamada no corpo da função principal

```
import pandas
def criar_df(entidades_gdt, entidades_hip):
  ent_PER_gdt = len(entidades_gdt[0])
  ent_ORG_gdt = len(entidades_gdt[1])
  ent_LOC_gdt = len(entidades_gdt[2])
  ent_MIS_gdt = len(entidades_gdt[3])
  ent_TOT_gdt = ent_PER_gdt + ent_ORG_gdt + ent_LOC_gdt + ent_MIS_gdt

  ent_PER_hip = len(entidades_hip[0])
  ent_ORG_hip = len(entidades_hip[1])
  ent_LOC_hip = len(entidades_hip[2])
  ent_MIS_hip = len(entidades_hip[3])
  ent_TOT_hip = ent_PER_hip + ent_ORG_hip + ent_LOC_hip + ent_MIS_hip

  ent_PER_pct = (ent_PER_gdt - ent_PER_hip)/(ent_PER_gdt + ent_PER_hip) \
                                                    if ent_PER_gdt else 0
  ent_ORG_pct = (ent_ORG_gdt - ent_ORG_hip)/(ent_ORG_gdt + ent_ORG_hip) \
                                                    if ent_ORG_gdt else 0
  ent_LOC_pct = (ent_LOC_gdt - ent_LOC_hip)/(ent_LOC_gdt + ent_LOC_hip) \
                                                    if ent_LOC_gdt else 0
  ent_MIS_pct = (ent_MIS_gdt - ent_MIS_hip)/(ent_MIS_gdt + ent_MIS_hip) \
                                                    if ent_MIS_gdt else 0
  ent_TOT_pct = (ent_TOT_gdt - ent_TOT_hip)/(ent_TOT_gdt + ent_TOT_hip) \
                                                    if ent_TOT_gdt else 0

  data = [[ent_TOT_hip, ent_PER_hip, ent_ORG_hip, ent_LOC_hip, ent_MIS_hip],
          [ent_TOT_gdt, ent_PER_gdt, ent_ORG_gdt, ent_LOC_gdt, ent_MIS_gdt],
          [ent_TOT_pct, ent_PER_pct, ent_ORG_pct, ent_LOC_pct, ent_MIS_pct]]

  nomes_colunas = ["TOT","PER","ORG","LOC", "MISC"]
  nomes_linhas = ["HIPÓTESE","GROUNDTRUTH","  NER"]
  df = pandas.DataFrame(data, nomes_linhas,nomes_colunas)
  return df
```

Exemplo de cálculo do NER

```
transcricao_gdt = """Sala de operação, Mickey, bom dia! Mickey, bom dia,
                     Donald aqui do COS. Oi, Donald. Ô Mickey, você
                     consegue verificar pra mim aí fazendo o favor, o montante
                     de geração que foi interrompido aí quando desligou aí o
                     trafo 3?"""

transcricao_hip = """Falha do coração Mickey Mickey Bom dia Donald aqui
                     do seu evento ou Mickey você consegue verificar para mim
                     fazendo favor o montante de geração que foi interrompido
                     aí desligou aí a outra preço."""

NER = calcular_metricas_NER(transcricao_gdt, transcricao_hip)
NER
```

---

**Groundtruth**

Sala de operação, Mickey PER , bom dia! Mickey PER , bom dia,

Donald PER aqui do COS ORG . Oi ORG , Donald PER . Ô Mickey PER , você consegue verificar pra mim aí fazendo o favor, o montante de geração que foi interrompido aí quando desligou aí o trafo 3?

---

**Hipótese**

Falha do coração Mickey Mickey Bom PER dia Donald PER aqui

do seu evento ou Mickey PER você consegue verificar para mim fazendo favor o montante de geração que foi interrompido aí

desligou aí a outra preço.

---

**Quantidade de entridades por tipo**

| | TOT | PER | ORG | LOC | MISC |
|---|---|---|---|---|---|
| **HIPÓTESE** | 3.0 | 3.00 | 0.0 | 0 | 0 |
| **GROUNDTRUTH** | 7.0 | 5.00 | 2.0 | 0 | 0 |
| **NER** | 0.4 | 0.25 | 1.0 | 0 | 0 |

# APÊNDICE H – PROCESSAMENTO DAS TRANSCRIÇÕES

```
def processar_transcricao(transcricao):
    print("Transcrição Original")
    print(transcricao)
    print("_" * 85)
    transcricao_processada = pre_processar_transcricao(transcricao)
    print("Transcrição Pré-Processada")
    print(transcricao_processada)
    print("_" * 85)
    transcricao_processada = remover_caracteres_especiais(transcricao_processada)
    print("Transcrição Sem Caracteres Especiais")
    print(transcricao_processada)
    print("_" * 85)
    transcricao_processada = remover_stopwords(transcricao_processada)
    print("Transcrição Sem Stop Words")
    print(transcricao_processada)
    print("_" * 85)
    transcricao_processada = aplicar_stemmer(transcricao_processada)
    print("Transcrição Com Aplicação de Stemming")
    print(transcricao_processada)
    print("_" * 85)

    return transcricao_processada
```

Exemplo de processamento das transcrições

```
transcricao_gdt = """Sala de operação, Mickey, bom dia! Mickey, bom dia,
                     Donald aqui do COS. Oi, Donald. Ô Mickey, você
                     consegue verificar pra mim aí fazendo o favor, o montante
                     de geração que foi interrompido aí quando desligou aí o
                     trafo 3?"""

transcricao_hip = """Falha do coração Mickey Mickey Bom dia Donald aqui
                     do seu evento ou Mickey você consegue verificar para mim
                     fazendo favor o montante de geração que foi interrompido
                     aí desligou aí a outra preço."""

transcricao_processada = processar_transcricao(transcricao_hip)
```

**Transcrição Original**
Falha do coração Mickey Mickey Bom dia Donald do seu evento ou Mickey você consegue verificar para mim fazendo favor o montante de geração que foi aí desligou aí a outra preço.

**Transcrição Pré-Processada**
falha do coração mickey mickey bom dia donald aqui do seu evento ou mickey você consegue verificar para mim fazendo favor o montante de geração que foi interrompido aí desligou aí a outra preço

**Transcrição Sem Caracteres Especiais**
falha do coracao mickey mickey bom dia donald aqui do seu evento ou mickey voce consegue verificar para mim fazendo favor o montante de geracao que foi interrompido ai desligou ai a outra preco

**Transcrição Sem Stop Words**
falha coracao mickey mickey bom dia donald aqui evento mickey voce consegue verificar mim fazendo favor montante geracao interrompido ai desligou ai outra preco

**Transcrição Com Aplicação de Stemming**
falh coraca mickey mickey bom dia donald aqu event mickey voc conseg verific mim faz favor mont geraca interromp ai deslig ai outr prec