INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Sistema de Transcrição Automática de Debates Parlamentares no Contexto da Assembleia da República Portuguesa

Pedro Miguel Delgado do Nascimento

Mestrado em Tecnologias Digitais para o Negócio

Orientador:
Professor Doutor João Carlos Amaro Ferreira,
Professor Auxiliar com Agregação,
ISCTE-IUL

Coorientador:
Professor Doutor Fernando Manuel Marques Batista,
Professor Associado,
ISCTE-IUL

Dezembro, 2023

TECNOLOGIAS
E ARQUITETURA

Departamento de Ciências e Tecnologias da Informação

Sistema de Transcrição Automática de Debates Parlamentares no Contexto da Assembleia da República Portuguesa

Pedro Miguel Delgado do Nascimento

Mestrado em Tecnologias Digitais para o Negócio

Orientador:
Professor Doutor João Carlos Amaro Ferreira,
Professor Auxiliar com Agregação,
ISCTE-IUL

Coorientador:
Professor Doutor Fernando Manuel Marques Batista,
Professor Associado,
ISCTE-IUL

Dezembro, 2023

Agradecimento

Gostaria de expressar a mais profunda gratidão à minha família, cujo apoio moral e incentivo incansável foram essenciais à realização deste mestrado. O vosso amor e encorajamento foram fundamentais a cada passo desta jornada.

Um agradecimento especial aos meus orientadores, Dr. João Carlos Ferreira e Dr. Fernando Batista, cuja visão e apoio constante foram essenciais na elaboração deste trabalho. A vossa orientação e experiência foram fundamentais, com conselhos valiosos e revisões críticas que enriqueceram significativamente este trabalho.

Estendo também os meus sinceros agradecimentos às minhas chefias e colegas na Assembleia da República. A vossa disposição em permitir que este trabalho incidisse sobre um processo da AR, o interesse que demonstraram na sua implementação, e o vosso auxílio na recolha de informações valiosas foram muito importantes para o sucesso deste projeto.

O meu sincero agradecimento a todos vós, por tornarem esta jornada não só possível, mas também profundamente enriquecedora. As vossas contribuições, apoio e confiança foram fundamentais para a conclusão do mestrado e serão sempre recordadas com grande apreço e gratidão.

Pedro Miguel Delgado do Nascimento

Resumo

A transcrição de debates parlamentares é essencial para a transparência, a responsabilidade e a acessibilidade da governação democrática. Tradicionalmente, esta tarefa é manual, morosa e depende de pessoas especializadas nessas tarefas. Este trabalho descreve o estudo, desenho e implementação de um sistema de transcrição para processamento automático dos debates parlamentares na Assembleia da República de Portugal, através de uma solução inovadora que utiliza tecnologias avançadas de reconhecimento automático de fala e de reconhecimento de mudança de orador.

O Sistema de Transcrição Automática (STAAR) foi desenvolvido após uma análise das tecnologias existentes e das necessidades específicas da Assembleia da República e resultou numa solução que se integra eficazmente com os processos existentes e prevê a evolução contínua da tecnologia. O sistema revelou-se de grande valor, não só pela sua eficácia e rapidez na transcrição de debates parlamentares, mas também pela sua capacidade em adaptar-se à linguagem parlamentar e ao jargão específico.

As conclusões deste trabalho revelam que o STAAR superou as expectativas ao apresentar uma taxa de erro bastante baixa, ao reduzir o tempo necessário para a produção do primeiro rascunho do Diário da AR e ao permitir a transcrição de reuniões de comissões parlamentares que anteriormente não eram documentadas, aumentando a abrangência e detalhe na documentação das atividades parlamentares.

Este avanço representa um passo significativo na modernização dos processos parlamentares, na promoção de maior transparência e acessibilidade das informações políticas e posiciona a Assembleia da República de Portugal na vanguarda da inovação tecnológica, no contexto da transcrição de debates parlamentares.

Palavras-chave: Transcrição automática, Debates parlamentares, Reconhecimento de fala, Processamento de linguagem natural, Aprendizagem automática, Diarização de orador

Abstract

The transcription of parliamentary debates is essential for transparency, accountability, and accessibility in democratic governance. Traditionally, this task is manual, time-consuming, and relies on specialized individuals. This work describes the study, design, and implementation of a transcription system for the automatic processing of parliamentary debates in the Assembly of the Republic of Portugal, through an innovative solution that uses advanced speech-to-text and speaker diarization technologies.

The Automatic Transcription System (STAAR) was developed after analyzing existing technologies and the specific needs of the Assembly of the Republic, resulting in a solution that effectively integrates with existing processes and anticipates continuous technological evolution. The system proved to be of great value, not only for its efficiency and speed in transcribing parliamentary debates but also for its ability to adapt to parliamentary language and specific terms.

The conclusions of this work show that STAAR exceeded expectations by presenting a very low error rate, reducing the time needed to produce the first draft of the Assembly of the Republic's Journal, and enabling the transcription of parliamentary committee meetings that weren't previously documented, thus increasing the scope and detail in the documentation of parliamentary activities.

This advancement represents a significant step in modernizing parliamentary processes, promoting greater transparency and accessibility of political information, and positions the Assembly of the Republic of Portugal at the forefront of technological innovation in the context of transcription of parliamentary debates.

Keywords: Automatic transcription, Parliamentary debates, Speech recognition, Natural language processing, Machine learning, Speaker diarization

Índice

Lista de Figuras

Lista de Tabelas

Lista de Acrónimos e Nomenclatura

AR	Assembleia da República
ASR	*Automatic Speech Recognition*
BPMN	*Business Process Model and Notation*
DAR	Diário da Assembleia da República
DER	*Diarization Error Rate*
DNN	*Deep Neural Network*
DSRM	*Design Science Research Methodology*
E2E	*End-to-end*
ECPRD	*European Centre for Parliamentary Research and Documentation*
GMM	*Gaussian Mixture Models*
GPU	*Graphics Processing Unit*
GRU	*Gated Recurrent Unit*
GT-TA	Grupo de trabalho da AR para a Transcrição Automática
HKT	*Hidden Markov toolkit*
HMM	*Hidden Markov Models*
IA	Inteligência Artificial
IEEE	*Institute of Electrical and Electronics Engineers*
LLM	*Large Language Model*
LTSM	*Long Short Term Memory*
NCHLT	*Nexus South African National Corpus Project*
NLP	*Natural Language Processing*
PRISMA	*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*
RNN	*Recurrent Neural Network*
SAE	*Standard American English*
STAAR	Sistema de Transcrição Automática da Assembleia da República
VAD	*Voice Activity Detection*
WER	*Word Error Rate*
WoSCC	*Web of Science Core Collection*

CAPÍTULO 1.

Introdução

1.1. Motivação

Com o avanço da tecnologia, as organizações e empresas têm cada vez mais acesso a ferramentas digitais que lhes permite melhorar processos e aumentar a eficiência do seu negócio. Nesse contexto, a utilização de tecnologias de reconhecimento de fala, tais como a conversão de fala em texto (*speech-to-text*), é vista como uma mais-valia em diversas áreas, incluindo na transcrição de debates parlamentares na Assembleia da República de Portugal (AR).

Desde a criação do registo escrito de debates parlamentares, a transcrição dos mesmos tem sido realizada manualmente por uma vasta equipa que tem a tarefa de transcrever todas as intervenções, apartes, aplausos e outros eventos sonoros que podem ocorrer durante uma sessão plenária e seja relevante documentar. Este processo de transcrição e produção de um diário legível e exato é demorado, uma vez que exige muito tempo e esforço humano, o que pode representar vários dias de trabalho até à produção e publicação do diário final.

A relevância da adoção de tecnologias que permitam realizar transcrições automáticas na AR foi enunciada no discurso da tomada de posse do Senhor Secretário-Geral da AR, Dr. Albino Azevedo Soares, no dia 19 de maio de 2022, que definiu a implementação de um sistema de transcrição automática dos trabalhos parlamentares como um dos grandes desafios da administração da AR. Nessa ocasião, o Senhor Secretário-Geral da AR acentuou a ideia de que a transcrição dos trabalhos parlamentares "é um trabalho técnico que necessita de enorme precisão e rigor" e que a implementação de um sistema de transcrição automática «corresponderá sem dúvida à preocupação de cada vez mais eficiência manifestada desde sempre (…) pela Divisão de Redação», incentivando o serviço a "trabalhar afincadamente no sentido da implementação deste sistema na Assembleia da República".

Com a utilização de tecnologias de *speech-to-text*, é possível automatizar esse processo, tornando-o mais rápido e eficiente. Atualmente existem tecnologias que permitem transcrever intervenções em tempo real, sem a necessidade de intervenção humana, ainda que, devido a diversos fatores, necessite de revisão humana para validação de eventuais erros na transcrição. Além de melhorar a eficiência do processo de transcrição, o uso de sistemas de reconhecimento de voz pode também contribuir para uma maior acessibilidade e inclusão, uma vez que ao disponibilizar as transcrições em tempo real ou pouco tempo após os debates parlamentares, pessoas com deficiência auditiva ou outras limitações poderão acompanhar e participar mais ativamente das discussões políticas.

1.2. Contexto

1.2.1. Dimensão Histórica

A transcrição de debates parlamentares em Portugal remonta ao início do século XIX, aquando da primeira reunião das Cortes Gerais e Extraordinárias da Nação Portuguesa a 24 de janeiro de 1821. Na época, as transcrições eram realizadas manualmente, através de estenografia e métodos taquigráficos, tendo este modelo de registo sido utilizado até aos anos 60 do século XX [1].

A partir da segunda metade dos anos 60, o Parlamento português adotou, como metodologia de registo, a gravação e transcrição integral das respetivas sessões plenárias. Com a introdução da gravação digital foram modificados por completo os modelos de registo.

É esse trabalho de mediação entre o discurso oral e o discurso escrito que fixa o texto que ficará a valer com valor político e histórico. A fixação de um texto com esta mediação exige, pois, grande concentração, rigor e um apurado conhecimento das posições políticas dos oradores e do momento em que decorrem os acontecimentos.

Foi apenas em 1992, com a criação do Diário da Assembleia da República em formato eletrónico, que a transcrição de debates parlamentares começou a ser informatizada. Esse formato permitiu que as transcrições dos debates fossem produzidas e disponibilizadas de forma mais rápida e eficiente.

Desde setembro de 2003 que as I e II Séries do Diário da Assembleia da República (DAR) são publicadas exclusivamente em formato eletrónico, disponibilizadas através do website do Parlamento e podem ser consultadas de forma gratuita e sem reservas por qualquer pessoa, garantindo a transparência e a documentação adequada dos processos democráticos em Portugal [2].

1.2.2. Evolução tecnológica

A transcrição é o processo de conversão da fala em texto escrito e desempenha um papel fundamental na comunicação humana e na preservação de informações. As tecnologias de transcrição têm passado por uma rápida evolução nos últimos anos, especialmente com o desenvolvimento da inteligência artificial e da aprendizagem profunda.

Antes de existirem tecnologias de reconhecimento de voz como atualmente as conhecemos, a transcrição manual era a única opção disponível. Indivíduos especializados nesta área ouvem áudios e transcrevem o conteúdo manualmente, um processo trabalhoso e demorado. Neste modelo a precisão e a qualidade da transcrição dependem muito das habilidades do transcritor e a revisão e edição são necessárias para garantir a qualidade do texto final.

Na década de 1970 surgiu a transcrição automática baseada em regras, que utilizava algoritmos programados para identificar palavras faladas. Estes sistemas eram limitados e funcionavam melhor com vocabulários pequenos e bem definidos, como em aplicações de reconhecimento de voz para comandos simples. No entanto, não eram suficientemente robustos para lidar com a complexidade e a variedade de discurso que se encontra em debates parlamentares [3].

No final da década de 1990, início dos anos 2000, a transcrição baseada em modelos de linguagem estatística começou a ganhar popularidade. Esses sistemas recorriam a algoritmos que analisavam grandes quantidades de dados para prever a probabilidade de palavras e sequências de palavras. Apesar das melhorias em relação aos sistemas baseados em regras, estes modelos ainda enfrentavam desafios significativos, como o reconhecimento de sotaques, dialetos e gírias [4].

A partir da década de 2010, as redes neuronais profundas e a inteligência artificial revolucionaram as tecnologias de transcrição. Esses sistemas são treinados com grandes volumes de dados, permitindo que aprendam padrões complexos e contextuais na linguagem falada. Esta abordagem melhorou significativamente a precisão e a velocidade da transcrição, tornando-a mais adequada para aplicações como transcrição de debates parlamentares [5].

1.2.3. Desafios na transcrição de debates parlamentares

Ao abordar a aplicação das tecnologias de transcrição aos debates parlamentares, é fundamental reconhecer os desafios específicos que esse contexto apresenta, nomeadamente:

- complexidade linguística - os debates parlamentares geralmente envolvem um alto nível de complexidade linguística, incluindo palavras técnicas, jargões e expressões idiomáticas;
- diferentes sotaques - os oradores de um debate podem ter diferentes sotaques e maneiras de pronunciar palavras, tornando mais difícil identificar corretamente as palavras faladas;
- velocidade da fala - os participantes num debate podem falar rapidamente, especialmente se estiverem emocionados ou tentando transmitir muitas informações num curto período;
- Interjeições e sobreposições - os debates são frequentemente caracterizados por interrupções, interjeições e sobreposições de falas, o que dificulta a transcrição clara e precisa do que está sendo dito, bem como a correta separação e identificação de vozes individuais;
- aplausos, ruídos e pausas - num debate, é comum ouvir aplausos, risos, expressões de apoio ou desacordo e outros ruídos de fundo;
- ambiguidade e referências - os participantes de um debate muitas vezes fazem referências a eventos, pessoas ou conceitos que não são explicados explicitamente, o que pode levar à ambiguidade e dificultar a compreensão do contexto. Essas referências podem ser

particularmente desafiadoras para transcritores caso não estejam familiarizados com o tema do debate;

- qualidade do áudio - a qualidade do áudio de um debate pode ser comprometida por problemas técnicos, como microfones de baixa qualidade ou interferência, que torna difícil a sua transcrição com precisão;
- emoção do discurso - os debates podem ser marcados por emoções intensas, paixões e ênfases que podem ser difíceis de capturar adequadamente através de sistemas automáticos de transcrição. A entoação, o sarcasmo e a ironia são elementos expressivos que enriquecem a comunicação e podem ser essenciais para a compreensão do conteúdo e do contexto do debate.

Outro desafio importante na implementação de tecnologias de *speech-to-text* é a adaptação dos atuais recursos humanos que fazem a transcrição e o impacto que essa mudança terá nas suas funções. É fundamental que as instituições, como a AR, considerem medidas para apoiar e requalificar os profissionais envolvidos no processo de transcrição manual.

Com a automação da transcrição, os transcritores humanos terão de se concentrar noutras atividades igualmente importantes, como rever e editar as transcrições geradas automaticamente, garantindo que elas se mantêm precisas e de alta qualidade. Além disso, poderão ser envolvidos no treino e aperfeiçoamento dos sistemas de reconhecimento de voz, utilizando sua experiência e conhecimento para melhorar o desempenho e a precisão dessas tecnologias.

1.3. Objetivos

Este trabalho de projeto de mestrado visa investigar e aplicar tecnologias de processamento de fala, em particular o reconhecimento automático de fala (*speech-to-text)* e o reconhecimento de mudança de orador, no contexto da transcrição de debates parlamentares, com o objetivo de melhorar o desempenho e a acessibilidade do processo de geração do Diário da Assembleia da República. Assim, os principais objetivos deste projeto são os seguintes:

1. Identificar e analisar diferentes tecnologias de *speech-to-text* e sua aplicabilidade na transcrição de debates parlamentares - este objetivo envolve a análise detalhada do estado da arte em transcrição automática de debates parlamentares, com foco especial na língua portuguesa, assim como avaliar experiências internacionais comparáveis;

2. Criar um sistema de transcrição automática no contexto da AR - com base nos resultados das análises e avaliações, bem como pelos problemas e requisitos recolhidos junto da equipa que atualmente produz manualmente as transcrições, o objetivo é desenvolver e implementar um sistema, com tecnologia *speech-to-text,* que possa ser utilizado pela AR no âmbito da transcrição de debates parlamentares. Este sistema deve ter em consideração questões como a integração da tecnologia nos processos existentes, a capacitação e requalificação dos transcritores humanos, a garantia da privacidade e segurança dos dados, e a manutenção e evolução contínua da tecnologia;
3. Implementar tecnologias de reconhecimento de mudança de orador para melhorar a transcrição de debates parlamentares - este objetivo tem por base a implementação de tecnologias diarização, que possam ser utilizadas em conjunto com as tecnologias de *speech-to-text* para diferenciar os diferentes participantes nos debates parlamentares, de forma a melhorar a qualidade e precisão das transcrições geradas automaticamente.

Com estes objetivos, este trabalho pretende contribuir para a pesquisa e prática na área de transcrição automática, aumentar a eficiência na transcrição de debates parlamentares, reduzir o tempo necessário para a tarefa de transcrição, realizar transcrições de reuniões que atualmente não têm registo escrito, bem como promover uma maior transparência e acessibilidade de informações políticas relevantes, contribuindo assim para um melhor entendimento e a abertura da Assembleia da República na sociedade.

1.4. Metodologia

O desenvolvimento deste projeto foi realizado seguindo a metodologia *Design Science Research Methodology* (DSRM) [6], uma abordagem que combina a teoria e a prática para resolver problemas específicos através da criação e avaliação de artefactos. A DSRM é caracterizada por um ciclo iterativo de pesquisa, que começa com a identificação e compreensão de um problema, seguido da conceção e construção de uma solução, e culmina na avaliação dessa solução no contexto real.

Como ilustrado na Figura 1, a DSRM é composta por várias etapas inter-relacionadas, que guiam o investigador desde a identificação do problema até à avaliação e refinamento da solução proposta.

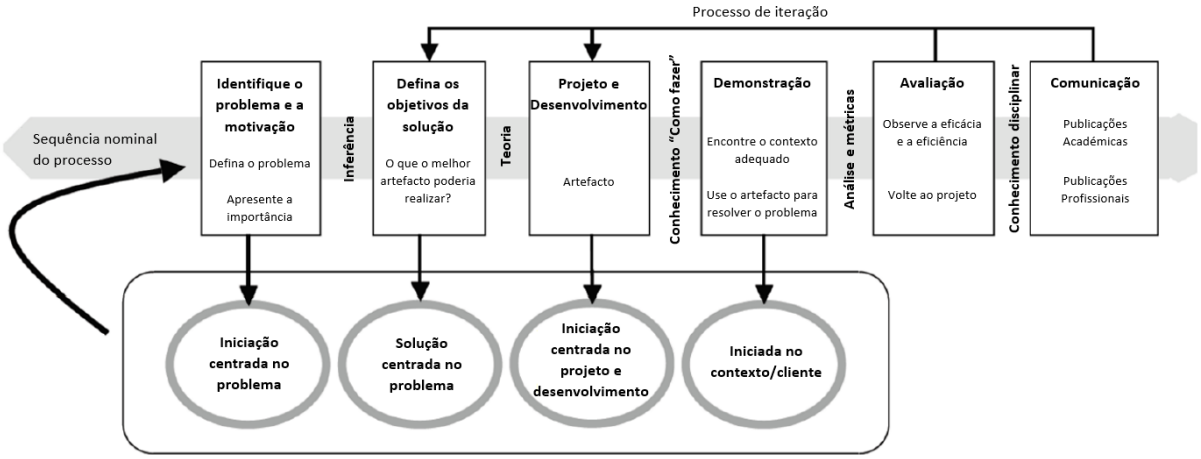


Figura 1 - Modelo de processo DSRM [6], adaptado para português

Dentro da metodologia DSRM, existem diferentes abordagens que podem ser adotadas, dependendo do foco e do contexto do projeto. Neste trabalho, optou-se pela abordagem de "Iniciação centrada no problema", que coloca ênfase na identificação e compreensão profunda do problema antes de avançar para as etapas de desenho e implementação.

O Sistema de Transcrição Automático da Assembleia da República (STAAR) foi concebido e desenvolvido seguindo as etapas da DSRM. Este processo foi executado em três iterações distintas, em que cada uma delas contribuiu para o aperfeiçoamento do sistema. Os detalhes destas iterações e das etapas específicas da DSRM que foram seguidas estão descritos no Capítulo 3 e Capítulo 4.

Para garantir que o desenvolvimento do projeto responde às necessidades identificadas (3.1), é fundamental avaliar o grau de cumprimento dos requisitos estabelecidos na primeira etapa do DSRM. A metodologia escolhida para esta avaliação é a escala NLPF da ISO 15504 [7], que se divide em quatro graus [8]:

- Não Atingido (N) - O requisito não foi cumprido ou foi atendido de forma muito limitada;
- Parcialmente Atingido (P) - O requisito foi em parte cumprido, mas ainda existem bastantes lacunas ou áreas que necessitam de melhorias;
- Largamente Atingido (L) - O requisito foi amplamente cumprido, com algumas áreas que precisam de melhoria;
- Totalmente Atingido (F) - O requisito foi completamente cumprido, sem lacunas identificadas.

Esta escala proporciona uma análise qualitativa precisa, permitindo uma visão clara de quais requisitos foram plenamente satisfeitos e quais ainda necessitam de ajustes ou melhorias.

No que se refere à avaliação da precisão de sistemas de reconhecimento de fala, uma das métricas mais comuns é a taxa de erro por palavra (WER, do inglês *Word Error Rate*). Esta métrica fornece uma quantificação objetiva da qualidade da transcrição, que permite comparar diferentes sistemas ou versões de um sistema e avaliar melhorias ao longo do tempo.

A WER é calculada através da comparação entre a transcrição gerada automaticamente e uma transcrição de referência, geralmente feita por um humano, de acordo com a Equação 1.

$$\text{WER} = \frac{\text{Substituições} + \text{Inserções} + \text{Eliminações}}{\text{Número total de palavras na transcrição de referência}}$$

Equação 1 - Cálculo da taxa de erro por palavra (WER)

Nesta fórmula, substituições refere-se ao número de palavras na transcrição automática que necessitam de ser alteradas para corresponder à transcrição de referência, inserções são palavras que aparecem na transcrição automática, mas não na transcrição de referência, e eliminações são palavras que estão na transcrição de referência, mas não aparecem na transcrição automática.

Ao longo das diversas etapas e iterações do processo de *Design Science Research Methodology* (DSRM), foi necessário quantificar a precisão das transcrições automáticas geradas. Para tal, recorreu-se a uma ferramenta *online*, disponibilizada pela AmberScript [9]. Esta ferramenta, concebida especificamente para esta finalidade, possibilita uma comparação direta entre as transcrições manuais, realizadas por profissionais e as transcrições produzidas por sistemas de transcrição automática, quantificando as diferenças de forma numérica através da *Word Error Rate* (WER). Assim, a cada iteração do DSRM, é possível avaliar de forma objetiva e quantitativa a eficácia e a precisão do sistema de transcrição automática desenvolvido.

1.5. Estrutura do documento

O presente trabalho está organizado em cinco capítulos e anexos. O Capítulo 1 apresenta o contexto geral do projeto, bem como os objetivos e motivação para a pesquisa. Estabelece-se ainda a importância do tema e esclarece-se sobre o que é abordado ao longo do trabalho. O Capítulo 2 é dedicado à revisão de literatura, onde são discutidas as pesquisas e trabalhos anteriores relacionados com o tema, inclusive noutros parlamentos que têm algum tipo de experiência nesta área. Esta parte do trabalho procura avaliar o estado atual do conhecimento na área, apresentando as principais teorias, abordagens e resultados alcançados. O Capítulo 3 apresentada uma visão detalhada sobre o sistema desenvolvido, desde a identificação dos requisitos até a sua implementação. O Capítulo 4

descreve o funcionamento do sistema desenvolvido e os casos onde foi utilizado, os resultados obtidos, bem como as principais dificuldades observadas e como foram ultrapassadas. Por último, o Capítulo 5 resume os principais resultados do projeto e discute as possibilidades de desenvolvimentos futuros, para reflexão sobre os avanços alcançados e algumas oportunidades de pesquisa que ainda podem ser explorados nesta área e tema em particular.

CAPÍTULO 2.

Trabalho Relacionado

A transcrição automática de fala para texto (*speech-to-text*) tem sido uma área de investigação com bastante interesse e de rápido desenvolvimento nos últimos anos. O acesso generalizado a tecnologias de reconhecimento de voz e os avanços nos algoritmos de processamento de linguagem natural permitiram grandes progressos nesta área. O presente capítulo apresenta uma análise detalhada do trabalho relacionado com o tema central deste trabalho de projeto, como forma de definir uma base teórica e contextual para a investigação.

Assim são introduzidos conceitos fundamentais, como os Sistemas de Reconhecimento Automático de Fala (ASR), e discutem-se as diversas abordagens existentes, com destaque à aplicação destes sistemas à língua portuguesa. É realizada uma análise dos sistemas de transcrição utilizados noutros parlamentos, o que permite perspetiva comparativa. A revisão de literatura abrange a metodologia de pesquisa adotada, os resultados encontrados e uma discussão crítica dos mesmos. São ainda exploradas tecnologias de reconhecimento de fala mais recentes, que representam avanços significativos no campo. A conclusão do capítulo sintetiza as informações relevantes obtidas e estabelece a importância destes para o desenvolvimento e avaliação de um sistema de transcrição automática para a Assembleia da República que constitui o foco deste trabalho.

2.1. Conceitos

Nesta secção, são apresentados os conceitos essenciais relacionados com reconhecimento automático de fala (ASR, do inglês *Automatic Speech Recognition*), os diferentes tipos e abordagens utilizados neste tipo de tecnologia, bem como suas aplicações no contexto da língua portuguesa.

Além disso, são abordadas as aplicações do ASR no contexto da língua portuguesa, apresentados exemplos de uso do ASR em diferentes áreas, como transcrição de áudio, assistentes virtuais, legendagem automática, sistemas de diálogo e muito mais e discutidas as contribuições e os desafios específicos relacionados à aplicação do ASR no idioma português, considerando suas peculiaridades fonéticas e linguísticas.

A compreensão destes conceitos é fundamental para o contexto da revisão de literatura, pois permite uma apreciação mais aprofundada dos estudos relacionados com ASR para a transcrição de debates parlamentares em língua portuguesa.

2.1.1. Sistemas de Reconhecimento Automático de Fala (ASR)

O reconhecimento automático de fala (ASR) refere-se a uma tecnologia que permite a transcrição automática da fala humana em texto escrito. Envolve o uso de algoritmos e modelos estatísticos para reconhecer e interpretar os sons da fala, convertendo-os em palavras e frases compreensíveis. Existem diferentes tipos e abordagens do ASR, incluindo modelos acústicos, modelos de linguagem e técnicas de aprendizagem de máquina. O ASR converte um sinal de fala numa representação textual, ou seja, uma sequência das palavras ditas, por meio de um algoritmo implementado com um módulo de software ou hardware. Existem vários tipos de fala natural, que pode ser classificada em [10]:

- fala soletrada (com pausas entre letras ou fonemas);
- fala isolada (com pausas entre palavras);
- fala contínua (quando um orador não faz pausas entre palavras);
- fala espontânea (num diálogo humano-humano);
- fala altamente conversacional (reuniões e discussões com várias pessoas).

Para diferentes tipos de fala, podem ser desenvolvidas soluções de ASR diferentes, e, portanto, os sistemas de ASR são geralmente construídos de acordo com o contexto e o objetivo.

2.1.2. Abordagens ASR

Desde o início dos anos 50 que existem diferentes abordagens para o desenvolvimento de sistemas ASR [11] e têm sido objeto de desenvolvimento contínuo até os dias de hoje. Ao longo dos anos, o foco principal dos sistemas ASR tem se deslocado entre diferentes tipos de abordagens, desde probabilísticas até os mais recentes modelos de redes neuronais profundas (DNN, do inglês *deep neural networks*) *end-to-end* (E2E, do inglês *End-to-End*), passando ainda por abordagens híbridas que combinaram as duas anteriores. Apesar da evolução do estado da arte, os sistemas ASR possuem componentes e processos que estão presentes na maioria das abordagens, como extratores de características, modelos acústicos e modelos linguísticos. De seguida estão enumeradas diferentes abordagens para desenvolver ASR através de diferentes métodos:

- Probabilísticos - Métodos como os modelos de Markov ocultos (HMM, do inglês *Hidden Markov Models*) têm sido uma abordagem popular em ASR durante muitos anos [12]. Neste método, os sinais de fala são modelados como uma sequência de estados e HMMs são utilizados para modelar as propriedades estatísticas dos sons da fala. Os sistemas ASR baseados em HMM utilizam uma combinação de modelos acústicos, para representar os sons da fala, e modelos de linguagem, para capturar o contexto linguístico. Estes modelos são treinados com grandes quantidades de dados de fala rotulados;

- Híbridos - Os sistemas ASR híbridos combinam diferentes modelos para aproveitar as suas melhores características. Como exemplo, uma abordagem híbrida comum é utilizar uma DNN para substituir o modelo acústico baseado em HMM num sistema tradicional, o que resulta num sistema ASR híbrido DNN-HMM [13]. Os modelos híbridos tiram partido tanto das características discriminativas das redes neuronais como das capacidades de modelagem linguística dos HMMs;
- End-to-End (E2E) com Redes Neuronais
    - Redes Neuronais Profundas (DNNs) - As técnicas de aprendizagem profunda, em particular redes neuronais profundas, têm avançado significativamente o ASR nos últimos anos [14]. Os sistemas ASR baseados em DNNs utilizam redes neuronais profundas para modelar a relação entre as características da fala e os fonemas. As redes são treinadas através de dados de fala rotulados e conseguem capturar padrões complexos nos dados, que resulta em melhorias na precisão de reconhecimento [15];
    - Redes Neuronais Recorrentes (RNNs) - As RNNs são um tipo de rede neuronal que consegue modelar dados sequenciais e torná-las adequadas para tarefas de ASR. Redes de Memória de Longo Prazo (LSTM) e Unidades Recorrentes com Portas (GRUs) são variantes populares de RNNs utilizadas em ASR [16] . As RNNs conseguem capturar dependências de longo alcance na fala e são frequentemente utilizadas em combinação com outros modelos, como HMMs ou DNNs [17];
    - Modelos baseados em Transformadores - Os *Transformers*, originalmente introduzidos para tarefas de processamento de linguagem natural, foram adaptados para ASR com um sucesso notável [18]. Os modelos baseados em *transformers* utilizam mecanismos de auto atenção para modelar as relações entre as características da fala e as transcrições. Estes modelos têm demonstrado um excelente desempenho, especialmente em tarefas que envolvem dependências de longo alcance e modelagem de contexto [19];

Estas abordagens podem variar em termos de técnicas subjacentes, metodologias de treino e requisitos computacionais. A pesquisa e desenvolvimento em ASR continua a evoluir com novas abordagens, como aprendizagem auto supervisionada [20] ou aprendizagem não supervisionada [21], que estão a ser exploradas para melhorar ainda mais o desempenho do ASR.

2.1.3. ASR para a língua portuguesa

A língua portuguesa, uma das mais faladas em todo o mundo, possui uma história rica e uma influência global significativa. O português, uma língua românica derivada do latim, é a língua oficial de nove países e seus habitantes, incluindo tanto a sua origem Portugal (10 milhões) como antigas colónias portuguesas: Brasil (216 milhões), Angola (33 milhões), Moçambique (32 milhões), Cabo Verde (570 mil), Guiné-Bissau (1,6 milhões), São Tomé e Príncipe (214 mil), Guiné Equatorial (13 milhões) e Timor-Leste (1,3 milhões) [22].

Apesar de ser a mesma língua, há variações linguísticas que diferem de país para país, mas também entre regiões do mesmo país, não só em termos de acentuação como de vocabulário, ainda que habitualmente inteligíveis, ou seja, que falantes destes países conseguem entender-se sem grandes dificuldades na comunicação.

Com base em pesquisas em bases de dados científicas foi possível identificar alguns estudos feitos sobre português de Portugal, também conhecido por português europeu. Identificam-se estudos relacionados com a dificuldade de expressão em pessoas muito jovens [23] ou mais idosas [24], pelo que procuraram, através de tecnologia ASR, melhor entendê-las.

Mais recentemente, em 2020, foi publicado um artigo [25] onde é feita uma revisão sobre ASR aplicado à língua portuguesa e suas variações, onde ou autores identificam que o português não é uma língua muito estudada nesta área. Realçam ainda os desafios em desenvolver ASR em português, fruto da necessidade de grandes quantidades de dados para uso em técnicas como redes neuronais, e como tal a necessidade de aumentar o estudo e explorar novas técnicas de classificação aplicadas a esta língua.

2.2. Sistemas de transcrição noutros parlamentos

Com recurso à rede do Centro Europeu de Pesquisa e Documentação Parlamentar (ECPRD) [26] foi enviado, em maio de 2022, um questionário a um conjunto de câmaras europeias, sobre experiência no uso de tecnologias de transcrição automática. Foram recebidas respostas dos parlamentos da Alemanha (Bundestag), Áustria (Nationalrat), Dinamarca, Espanha (Congreso de los Diputados), Estónia, Finlândia, França (Sénat), Grécia, Irlanda e Países Baixos (Tweede Kamer der Staten-Generaal).

Face às respostas recebidas, obteve-se a seguinte informação relevante para a implementação de um sistema de transcrição automática na AR:

- o número de pessoas envolvidas na transcrição do jornal oficial varia entre 19 e 65;
- a maioria das equipas transcreve tanto sessões de Plenário como reuniões de Comissões;
- o número de horas de transcrição por semana varia entre 17 horas e 80 horas;

- na maioria dos parlamentos existem três níveis de revisão, transcrição /revisão / edição ou publicação final, sendo disponibilizada uma versão inicial, sem revisão final, num período que pode variar desde o final da reunião até 48 horas depois da reunião;
- o prazo de publicação final varia entre o dia seguinte até um ano depois da data da reunião;
- há casos em que o software de transcrição foi desenvolvido especificamente para uso pelo parlamento, sempre em parceria com o serviço responsável pelas tecnologias de informação, mas a maioria usa sistemas disponíveis no mercado, que implicam que o texto seja ditado para o sistema;
- o uso do software de transcrição automática fica ao critério de cada redator, havendo parlamentos em que quase todos os redatores o utilizam, embora alguns prefiram usar um ficheiro de áudio e o PC para fazer a transcrição, dependendo especialmente do orador a transcrever;
- todos os parlamentos responderam que o texto resultante da transcrição automática nunca poderia ser publicado diretamente, sendo necessários vários níveis de intervenção, uma vez que, segundo algumas respostas, os programas não reconhecem determinadas estruturas gramaticais e frásicas e nomes estrangeiros, além de não registarem aplausos e apartes e de ser sempre necessário verificar siglas e erros de compreensão.

Foram assinaladas como vantagens a questão da preservação da saúde, com menor cansaço físico e menor probabilidade, a médio/longo prazo, de doenças profissionais devido ao uso intensivo e prolongado do teclado, e a possibilidade de teletrabalho, com um ganho de tempo na fase de digitação do texto. Já como desvantagens foram identificadas a atenção redobrada que é necessária na revisão, devido à má qualidade da transcrição automática e o consequente aumento do tempo despendido na fase de revisão final para manter os padrões de qualidade. Além de não dispensar a inclusão manual de fórmulas de votação, de apartes, dos nomes dos oradores, de fórmulas regimentais, entre outras, a qualidade do texto transcrito automaticamente depende muito da dicção do orador e do tipo de debate.

A experiência obtida a partir das informações de outros parlamentos desempenha um papel fundamental na criação de um sistema de transcrição automática para o parlamento português. Ao conhecer as vantagens e desvantagens relatadas por outras câmaras, é possível adaptar e otimizar a implementação dessa tecnologia, tendo em consideração as necessidades e especificidades do contexto da AR. Isto permite uma abordagem mais informada e eficiente na construção de um sistema que tenha em conta as necessidades específicas da Assembleia da República.

2.3. Revisão de literatura

Como forma de ter uma visão abrangente do estado atual do conhecimento sobre uma determinada área de estudo, a revisão de literatura desempenha um papel vital e envolve a pesquisa, análise e síntese de estudos existentes sobre esse tópico.

Para a análise do estado de arte sobre *speech-to-text* foi realizada uma revisão de literatura através da metodologia PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) [27]. A PRISMA é uma metodologia amplamente aceite e recomendada para a realização de revisões sistemáticas, uma vez que fornece orientações claras e rigorosas para o planeamento, execução e análise desses estudos. A adoção do PRISMA permite garantir a transparência, reprodutibilidade e qualidade do processo de revisão sistemática.

2.3.1. Metodologia de pesquisa

A estratégia de pesquisa foi realizada de acordo com as diretrizes da PRISMA [27] o que resultou num processo metódico, passo a passo. Foram efetuadas pesquisas nas bases de dados *Scopus* [28], IEEE Xplore [29] e *Web of Science Core Collection* (WoSCC) [30], com base em termos de pesquisa apropriados para identificar estudos de interesse no campo de reconhecimento automático de fala no âmbito de transcrição de debates parlamentares, que têm particularidades distintas de outro tipo de discurso.

A seleção destas bases de dados foi feita tendo em conta a sua ampla cobertura de literatura científica e técnica. Utilizaram-se termos de busca apropriados, especificamente elaborados para abranger o âmbito da pesquisa. Os termos a procurar foram definidos tendo em conta palavras-chave relativos a três dimensões, a dimensão "tecnologia", relacionada com reconhecimento automático de fala, a dimensão "propósito", que envolve transcrição e diarização e a dimensão "âmbito", relacionada com as particularidades do domínio específico de debates parlamentares.

Assim, os repositórios foram pesquisados sistematicamente em relação a trabalhos publicados relativos a estas três dimensões, tecnologia, propósito e âmbito.

Para garantir a atualidade dos estudos incluídos na revisão, foram considerados apenas resultados publicados desde 2013 até maio de 2023, abordagem que permitiu abranger a literatura mais recente e relevante neste campo de estudo.

Foi estabelecido ainda o critério de seleção de estudos escritos em português ou inglês, face à disponibilidade e predominância da literatura científica nesses idiomas no contexto da pesquisa em reconhecimento automático de fala.

A Tabela 1 sumariza os termos que foram usados na pesquisa bem como os filtros aplicados.

Tabela 1 - Termos pesquisados e filtros aplicados

Tecnologia	Propósito	Âmbito	Filtros
ASR NLP Speech-To-Text Speech Recognition	Transcription Diarization	Parliament*	Apenas publicados em português ou inglês, entre 2013 e 2023

Com base nas três dimensões e termos identificados, foi criada uma *string* de pesquisa que foi usada nas bases de dados de conhecimento consideradas:

((asr OR nlp OR speech-to-text or "speech recognition")
AND (transcription OR diarization) AND (parliament))*

A seleção inicial de publicações foi efetuada, de acordo com a metodologia PRISMA [27], através da análise dos títulos e resumo por forma a avaliar a relevância e adequação ao tema em estudo. Posteriormente foi verificado se o texto completo dos artigos estava disponível para consulta, incluindo na revisão apenas aqueles em é possível efetuar essa análise.

Na revisão da literatura, para além dos artigos científicos selecionados através de bases de dados académicas, também foram incluídas outras fontes de conhecimento, como recursos online relevantes e citações provenientes de especialistas na área.

Posteriormente foi realizada a leitura integral dos documentos resultantes da triagem para avaliar a sua adequação aos critérios de inclusão.

Os resultados foram guardados na ferramenta Zotero [31] e tratados estatisticamente com o Microsoft Excel [32], nomeadamente através de dados como título, autor, ano, publicação, área, palavras-chave e resumo.

2.3.2. Resultados

De acordo com a metodologia PRISMA, seguindo as premissas estabelecidas em relação às palavras-chave, ao ano de publicação e ao idioma, foi conduzida uma pesquisa nas bases de dados selecionadas. Inicialmente, foram identificados 60 registos que cumprem os critérios de inclusão estabelecidos. Estes registos foram submetidos a uma análise preliminar para determinar sua relevância e adequação à revisão sistemática em questão.

A Tabela 2 descreve e quantifica os artigos que cada dimensão, e suas *keywords,* retornaram ao ser pesquisadas nas bases de dados científicas consultadas, bem como o resultado da filtragem a cada dimensão incluída.

Tabela 2 - Termos pesquisados e resultados por base de dados

<table>
<tr><th>Tecnologia</th><th>Propósito</th><th>Âmbito</th><th>Filtros</th></tr>
<tr><td>ASR
NLP
Speech-To-Text
Speech Recognition</td><td>Transcription
Diarization</td><td>Parliament*</td><td rowspan="4">Publicados em português ou inglês, entre janeiro de 2013 e maio de 2023</td></tr>
<tr><td>136,042 registos
(Scopus - 133,109)
(WoS - 62,714)
(IEEE - 43,130)</td><td>2,110,670 registos
(Scopus - 1,369,019)
(WoS - 736,731)
(IEEE - 4,920)</td><td>88,804 registos
(Scopus - 54,557)
(WoS - 33,753)
(IEEE - 494)</td></tr>
<tr><td colspan="2">7,865 registos
(Scopus - 4,932)
(WoS - 1,304)
(IEEE - 1,629)</td><td></td></tr>
<tr><td colspan="3">130 registos
(Scopus - 67)
(WoS - 44)
(IEEE - 19)</td></tr>
<tr><td colspan="4">60 registos
(Scopus - 33)
(WoS - 20)
(IEEE - 7)</td></tr>
</table>

Após a identificação de artigos que têm os termos das três dimensões analisadas, de acordo com a metodologia descrita no PRISMA [27], verificou-se que 25 eram duplicados e como tal, foram removidos do conjunto de estudos.

Em seguida, os títulos e resumos dos 35 registos restantes foram analisados. Nesse processo de triagem, 13 registos não foram considerados relevantes para o objeto de estudo, convergindo em 22.

Foi feita uma pesquisa adicional para verificar a disponibilidade dos estudos selecionados. Nesse sentido, 17 publicações não estavam disponíveis para leitura completa, sendo por isso excluídos do conjunto de estudos em análise, o que resultou em 5 publicações para análise mais detalhada.

Foram ainda incluídos 4 artigos obtidos através de pesquisas em *websites*, cujo conteúdo mostrou-se potencialmente de relevância para o tema em estudo.

Após a fase de triagem e seleção de registos, restaram no final do processo 9 estudos. Estas publicações representam o conjunto final de estudos que foram analisados e discutidos na revisão de literatura, com o objetivo de fornecer uma base sólida para a compreensão do estado atual do conhecimento sobre o tema em questão.

A Figura 2 apresenta o fluxograma do processo metodológico adotado na revisão sistemática, seguindo as diretrizes da metodologia PRISMA [27], como forma de demonstrar as etapas seguidas desde a identificação inicial dos estudos até a seleção final dos artigos incluídos na revisão.

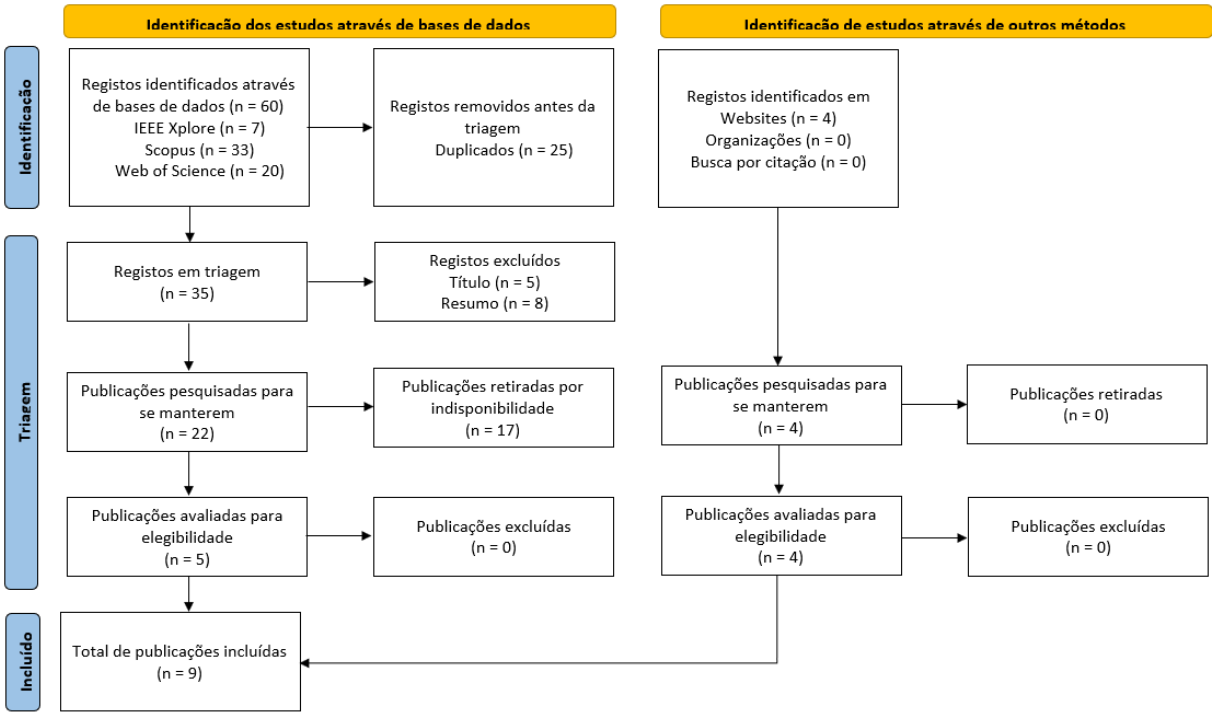


Figura 2 - Fluxograma PRISMA para revisão sistemática

2.3.3. Discussão

Nesta secção é feita a análise em resumo dos artigos identificados através da metodologia PRISMA, com o objetivo de explorar e interpretar os principais resultados através revisão sistemática realizada. Com base nesta análise foi criada a Tabela 3 que sintetiza os pontos mais relevantes de cada um destes artigos, contribuindo assim para uma maior compreensão do tema de transcrição automática e uma base sólida para este trabalho.

Tabela 3 - Análise de estudos incluídos na revisão sistemática

Publicação	Tecnologia/ferramenta	Abordagem*			Dados	Idioma	Propósito	Erro (%)
		P	H	RN				
Zhao et al. (2023), A Survey of Large Language Models [33]	Modelo de linguagem de grande dimensão (LLM)			✓✓		Multilingue		
Vos et al. (2023), Political corpus creation through automatic speech recognition on EU debates [34]	Transformador wav2vec2.0			✓✓	56.300h	Inglês	Transcrição	17,9
Dıáz-Munío et al. (2021), Europarl-ASR: A large corpus of parliamentary debates for streaming ASR benchmarking and speech data filtering/verbatimization [35]	4-gram Transformador FairSeq	✓✓		✓✓	1.300h	Inglês	Transcrição	7,0 ~ 7,9
Lima et al. (2020), A survey on Automatic Recognition systems for Portuguese language and its variations [25]	Máquinas de vetores de suporte (SVM) Modelos Markov ocultos (HMM) Redes neuronais convulsionais (CNN) Redes neuronais profundas (DNN)	✓✓	✓✓	✓✓		Português		
Alumaë et al. (2018), Advanced rich transcription system for Estonian speech [36]	Kaldi *Toolkit* LIUM SpkDiarization *toolkit*		✓✓	✓✓	690h	Estónio	Transcrição e Diarização	8,1
Kawahara (2018), Automatic meeting transcription system for the Japanese parliament (diet) [37]	Tradução automática estatística (SMT) Treino HMM baseado em citéria de máxima semelhança (*Maximum Likelihood*)	✓✓		✓✓	200h	Japonês	Transcrição	10,0
Mansikkaniemi et al. (2017), Automatic construction of the Finnish parliament speech corpus [38]	Algoritmo Levenshtein Modelos acústicos DNN Alinhamento *speech-to-text*		✓✓		2.000h	Finlandês	Transcrição	5,9 ~ 18,7
De Wet et al. (2016), Developing Speech Resources from Parliamentary Data for South African English [39]	*Hidden Markov toolkit* (HTK) Dicionários pronúncia NCHLT, SAE	✓✓			105h	Inglês África do Sul	Transcrição	
Campr et al. (2014), Audio-video speaker diarization for unsupervised speaker and face model creation [40]	Modelos de Mistura Gaussiana (GMMs) Deteção de atividade de voz (VAD) Algoritmo maximização expetativa (*EM)*	✓✓			30h	Checo	Diarização	7,2

* Probabilística (P), Híbrida (H), Redes neuronais (RN)

A evolução da transcrição automática, em particular no contexto dos debates parlamentares, tem sido marcada por avanços tecnológicos significativos e mudanças paradigmáticas. Inicialmente, os sistemas de reconhecimento automático de fala (ASR) baseavam-se em modelos probabilísticos. Estes modelos, como os modelos de mistura Gaussiana (GMMs) e os modelos Markov ocultos (HMMs), são fundamentados em estatísticas e probabilidades.

No âmbito da transcrição de discursos, o estudo de De Wet et al. [39], publicado em 2016, descreve o desenvolvimento de recursos de fala específicos para o inglês sul-africano, uma língua com bastantes nuances e variações que apresentam desafios únicos para os sistemas de reconhecimento automático de fala (ASR). Face a estes desafios, os autores focaram-se na adaptação do ASR para capturar com precisão as peculiaridades desse dialeto. Para isso, utilizaram o *Hidden Markov toolkit* (HTK) [41], uma ferramenta bastante completa que permite modelar sequências, como séries temporais de fala.

Além do HTK, a pesquisa também se baseou em dicionários de pronúncia, especificamente criados para o inglês sul-africano (NCHLT e SAE), que forneceram o suporte linguístico necessário para treinar e refinar o modelo ASR. A combinação dessas ferramentas e recursos resultou em melhorias significativas na precisão da transcrição, demonstrando a eficácia de abordagens especializadas em ambientes linguísticos específicos.

Ao melhorar a deteção e transcrição da fala em inglês sul-africano, De Wet et al. [39] demonstraram que, com as ferramentas e recursos corretos, é possível adaptar sistemas ASR a diversas línguas e dialetos.

A introdução de modelos híbridos veio dar resposta aos desafios dos modelos puramente probabilísticos. Em 2017, Mansikkaniemi et al. [38] focaram-se na construção automática de um *corpus* de discursos do parlamento finlandês. Os autores enfrentaram vários desafios durante a criação do *corpus*, incluindo o acesso limitado a dados de treino de domínio geral em grande escala e a necessidade de grandes quantidades de dados de fala transcritos. Através da combinação de modelos probabilísticos, como algoritmo *Levenshtein,* e modelos acústicos baseados em redes neuronais profundas (DNN), o estudo permitiu criar um *corpus* de transcrição a partir dos dados disponíveis no portal web do Parlamento da Finlândia.

Após uma utilização vasta dos modelos híbridos, a evolução tecnológica e os avanços na pesquisa levaram ao aparecimento das redes neuronais *end-to-end, que* prometiam simplificar a arquitetura dos sistemas ASR, ao eliminar a necessidade de múltiplas etapas de processamento e permitir uma abordagem mais direta e integrada à transcrição automática. As redes neuronais, com a sua capacidade de aprender padrões complexos, começaram a dominar o campo do ASR.

Em 2018, Alumaë et al. [36] publicaram um artigo onde é descrito o desenvolvimento de um sistema avançado de transcrição para o idioma estónio com base no *Kaldi toolkit* [42], um software livre para reconhecimento de fala que suporta diversos tipos de redes neuronais. Um dos problemas enfrentados pelo sistema foi lidar com dados gravados "em ambientes naturais", como entrevistas e reuniões gravadas em condições acústicas adversas, habituais no mundo real. Para superar esses problemas, o sistema foi treinado com uma variedade diversificada de dados de fala e foram utilizadas técnicas como redução de ruído para melhorar os resultados. O estudo demonstrou a eficácia das redes neuronais na transcrição, que alcançou resultados notáveis em termos de precisão, com uma taxa de erro WER de 8,1%. De acordo com os autores, o sistema desenvolvido tem potencial para se adaptar a outras línguas uma vez que utiliza uma combinação de modelos acústicos e linguísticos que podem ser treinados com dados de fala que estejam noutros idiomas.

Também em 2018, Kawahara [37] focou-se na transcrição automática de reuniões do parlamento japonês, mas com uma abordagem diferente da que foi utilizada por Alumaë et al. [36], referida anteriormente. Kawahara [37] utilizou métodos probabilísticos, como a tradução automática estatística (SMT) e treino HMM baseado em critérios de máxima semelhança (ML), bem como explorou as capacidades das redes neuronais, na perspetiva de demonstrar como a combinação de diferentes técnicas podem levar a resultados superiores. O sistema ASR foi desenvolvido com uma combinação de software de código aberto e proprietário, e foi treinado com grande *corpus* de dados de áudio de reuniões parlamentares anteriores. O sistema foi avaliado através da operação a longo prazo em ambiente do parlamento japonês, e registou uma taxa de precisão de cerca de 90%. Um dos principais desafios enfrentados pelo autor foi a variabilidade da fala nas reuniões parlamentares, que inclui disfluências, preenchimentos e expressões coloquiais. Para enfrentar esse desafio foi adotada, de acordo com o autor, uma abordagem sustentável que combina a transcrição automática e edição manual como forma de gerar uma transcrição fiel da reunião.

A nível europeu, em 2021 Dıáz-Munío et al. [35] deram um passo significativo na área de reconhecimento automático de fala (ASR) ao criar o Europarl-ASR, um *corpus* notável não só pelo seu tamanho, com 1.300 horas de discursos parlamentares transcritos, mas também pela sua diversidade linguística, refletindo as diversas línguas faladas no Parlamento Europeu [43].

Ao trabalhar com um *corpus* tão diversificado, Dıáz-Munío et al. [35] também enfrentaram desafios relacionados com a grande variedade de estilos de fala, sotaques, nuances linguísticas e com terminologia técnica e específica, bem como referências culturais e históricas que podem variar de uma língua para outra. Para superar estas questões, os autores utilizaram uma combinação de redes neuronais profundas (DNN) e redes neuronais recorrentes (RNN). As DNNs, com a sua capacidade de

aprender representações hierárquicas de dados, foram fundamentais para capturar as características acústicas da fala. Por outro lado, as RNNs, com a sua habilidade de modelar sequências temporais, foram essenciais para entender a estrutura e a sequência dos discursos.

Os resultados do estudo foram promissores, onde se obteve taxas de erro entre os 7 e 7,9%, o que demonstrou que as redes neuronais não só são capazes de transcrever debates parlamentares com precisão, mas também de se adaptar e generalizar para várias línguas, embora o português não tenha sido avaliado nem esteja previsto nos planos futuros, de acordo com as conclusões apresentadas.

Recentemente, em 2023, o Europarl-ASR foi utilizado como modelo de treino e avaliação num estudo realizado por Vos et al. [34], que criou um *corpus* com transcrições do comité LIBE do Parlamento Europeu [44], com um total de 3,6 milhões de palavras.

O foco principal do estudo foi explorar e otimizar técnicas avançadas de redes neuronais para ASR. Em vez de se apoiarem exclusivamente no *Kaldi toolkit*, como estudos anteriores, os autores incorporaram o modelo transformador wav2vec2.0 [20] na sua *pipeline* de ASR. Este modelo, conhecido pela sua capacidade de aprender representações ricas de áudio de forma não supervisionada, permitiu uma abordagem inovadora para transcrever debates parlamentares. A escolha do wav2vec 2.0 foi estratégica, uma vez que dada a complexidade e a diversidade linguística dos debates da UE, era essencial utilizar uma ferramenta capaz de capturar nuances subtis na fala. No entanto, mesmo com a incorporação do wav2vec 2.0, a taxa de erro de palavra (WER) obtida foi de 14,5% o que indica que, embora o modelo seja eficaz, ainda há espaço para otimização.

Os artigos encontrados nesta revisão sistemática, que descrevem a utilização de sistemas de ASR, não mencionam especificamente a transcrição no idioma português, que possui particularidades únicas. Neste contexto, o trabalho de Lima et al. [25], publicado em 2020, oferece uma perspetiva abrangente sobre o estado atual dos sistemas ASR para este idioma.

O estudo consistiu numa revisão sistemática de 101 artigos publicados entre 2012 e 2018, através da metodologia PRISMA, com o objetivo de identificar as técnicas de ASR mais usadas, bem como avaliar os desafios enfrentados no desenvolvimento de sistemas ASR para o português, como a falta de recursos e a alta variabilidade da língua. Os autores verificaram que a maioria das pesquisas científicas sobre ASR para o português concentram-se no português europeu, com muito pouco trabalho realizado sobre o português brasileiro ou outras variações da língua. Observam ainda que, enquanto alguns artigos têm como objetivo recolher dados de fala em português, a maioria deles não os publica de forma gratuita, o que torna difícil utilizar técnicas mais sofisticadas, como DNNs e CNNs, que requerem uma grande quantidade de dados. Os autores concluem que apesar das redes neuronais profundas (DNN) se terem tornado a técnica mais comum para ASR nos últimos anos, alguns métodos

inovadores para ASR em português estão por explorar, como a aprendizagem por transferência ou aprendizagem não supervisionada.

Outro estudo particularmente interessante e atual é o de Zhao et al. [33], publicado em 2023, que consiste numa revisão literária sistematizada sobre técnicas e aplicações inovadoras com modelos de linguagem de grande dimensão (LLM). A pesquisa destaca a importância do dimensionamento de modelos para alcançar o máximo desempenho em tarefas de processamento de linguagem natural. Os autores avaliaram os desafios atuais no dimensionamento de modelos linguísticos, incluindo recursos computacionais e disponibilidade de dados que é sempre o mais difícil de obter, e como esses desafios podem ser superados através de técnicas de treino distribuídos e aumento de dados. Em termos de tecnologia e modelos de linguagem avaliados, o estudo incide sobre um grande leque de modelos como BERT [45], RoBERTa [46], T5 [47] e GPT [48]. Em relação a este último modelo, os autores reconhecem o nível de excelência do GPT-4 na resolução de tarefas gerais e sua robustez contra ruídos ou perturbações. Resultados empíricos do estudo mostram que o GPT-4 supera os restantes modelos avaliados numa ampla variedade de tarefas, como a compreensão de linguagem.

Uma das conclusões que se pode ler na publicação é que os modelos de linguagem de grande dimensão revolucionaram o campo de processamento de linguagem natural e têm o potencial de transformar muitos outros campos, incluindo pesquisa científica, saúde e educação, ainda que com reservas no que diz respeito a questões éticas e de segurança em torno da AI.

Um dos objetivos deste trabalho de tese é complementar a transcrição dos debates parlamentares com a identificação de mudança de orador. Este processo, chamado diarização, consiste em distinguir e segmentar uma gravação de áudio ou vídeo para determinar "quem falou quando". Nesse sentido, os estudos de Campr et al. [40], de 2014, e de Alumaë et al. [36], de 2018, são relevantes, uma vez que fornecem metodologias e experiência que podem ser aplicadas ou adaptadas para melhorar a precisão e eficácia da diarização no contexto de debates parlamentares.

Campr et al. [40] utilizaram modelos de mistura Gaussiana (GMMs) para abordar a tarefa de diarização. GMMs são modelos probabilísticos que podem ser treinados para reconhecer e distinguir diferentes oradores com base nas características únicas da sua forma de falar. Ao combinar isso com algoritmos de maximização de expectativa (EM), os autores foram capazes de segmentar eficazmente as gravações e atribuir segmentos a oradores individuais. Um dos principais desafios enfrentados foi a deteção de atividade de voz em ambientes parlamentares, que podem ter bastante ruído de fundo, aplausos, interjeições e outras interrupções. Para minimizar estas situações, o sistema utiliza uma combinação de áudio e de vídeo, bem como tecnologia de reconhecimento facial, para associar

modelos individuais das modalidades de áudio e vídeo de forma não supervisionada. O sistema foi avaliado em 30 horas de vídeo, especificamente em transmissões de reuniões do parlamento checo. Os resultados mostram que a combinação proposta dos sistemas individuais de diarização de áudio e vídeo resulta numa melhoria da taxa de erro de diarização (DER), avaliada em 7,2%.

Já a investigação de Alumaë et al. [36], publicada em 2018, utiliza apenas o áudio para realizar a diarização, que é efetuada através do *LIUM SpkDiarization toolkit* [49]. Esta ferramenta de código aberto utiliza uma combinação de técnicas de agrupamento e classificação para identificar oradores em segmentos de áudio. De acordo com os autores, através desta ferramenta foi possível atingir uma taxa de precisão de 95% no processo de diarização.

2.4. Tecnologias de reconhecimento de fala

No cenário atual da ciência da computação e da linguística computacional, o reconhecimento automático de fala tem sido uma área de pesquisa e desenvolvimento em rápido crescimento. Diversas tecnologias têm sido propostas e melhoradas ao longo dos anos, com o objetivo de obter transcrições precisas e eficientes em diferentes contextos e línguas. Entre essas inovações, o Whisper [50], um modelo de linguagem de grande dimensão (LLM), surgiu com grande potencial não só pelo volume impressionante de dados com os quais foi treinado, mas também pela sua versatilidade e precisão em transcrever múltiplas línguas, incluindo o português.

Nesta secção, pretende-se detalhar as especificidades do Whisper, entender outros produtos derivados e entender as motivações e critérios que levaram à sua seleção como a tecnologia central para a criação do Sistema de Transcrição Automática da Assembleia da República (STAAR).

2.4.1. Whisper

Em setembro de 2022, a OpenAI, a empresa por trás do ChatGPT [51] e do DALL-E [52], disponibilizou o seu novo modelo de reconhecimento automático da fala (ASR), o Whisper [50], um sistema multilingue e multitarefa que tem um desempenho cada vez mais próximo do nível humano. O Whisper dispõe de documentação a detalhar várias ideias e técnicas novas e interessantes por trás da sua estratégia de formação e construção de conjuntos de dados, que lhe permite ter um desempenho bastante bom em diversas línguas e sotaques. De forma única, o Whisper suporta a transcrição em várias línguas, bem como a tradução dessas línguas para inglês. A OpenAI detalha que o modelo foi treinado em 680.000 horas de dados supervisionados, equivalente a mais de 77 anos de áudio contínuo. De acordo com o estudo publicado em 2022 por Radford et al. [53], o Whisper atinge uma robustez e precisão de nível humano quando executado em discurso em inglês.

Para a arquitetura do modelo, o OpenAI utilizou um modelo de codificador-descodificador baseado em transformadores bastante simples. Introduzida por Vaswani et al. [54], em 2017, esta arquitetura tem sido responsável pela maioria das recentes descobertas no campo do processamento da linguagem natural e ASR.

A abordagem passa por dividir o áudio de entrada em blocos curtos de 30 segundos, que são convertidos num bloco de espetrograma log-Mel e passam depois para um bloco codificador, onde o espetrograma é processado. O descodificador encarrega-se de várias tarefas por si só, tais como a identificação da língua, os carimbos temporais ao nível da frase, a transcrição do discurso multilingue e a tradução do discurso para inglês. Através um conjunto personalizado de *tokens* especiais, o descodificador decide em que tarefa específica deve trabalhar. O descodificador, sendo um modelo de linguagem áudio-condicional, é também treinado para utilizar o histórico do texto da transcrição com o objetivo de lhe fornecer algum contexto e ajudar a lidar com áudio ambíguo.

A Figura 3 representa um diagrama de funcionamento do Whisper [55] , adaptado para português, onde podem ser observadas as diversas fases no processo de transcrição descrito anteriormente.

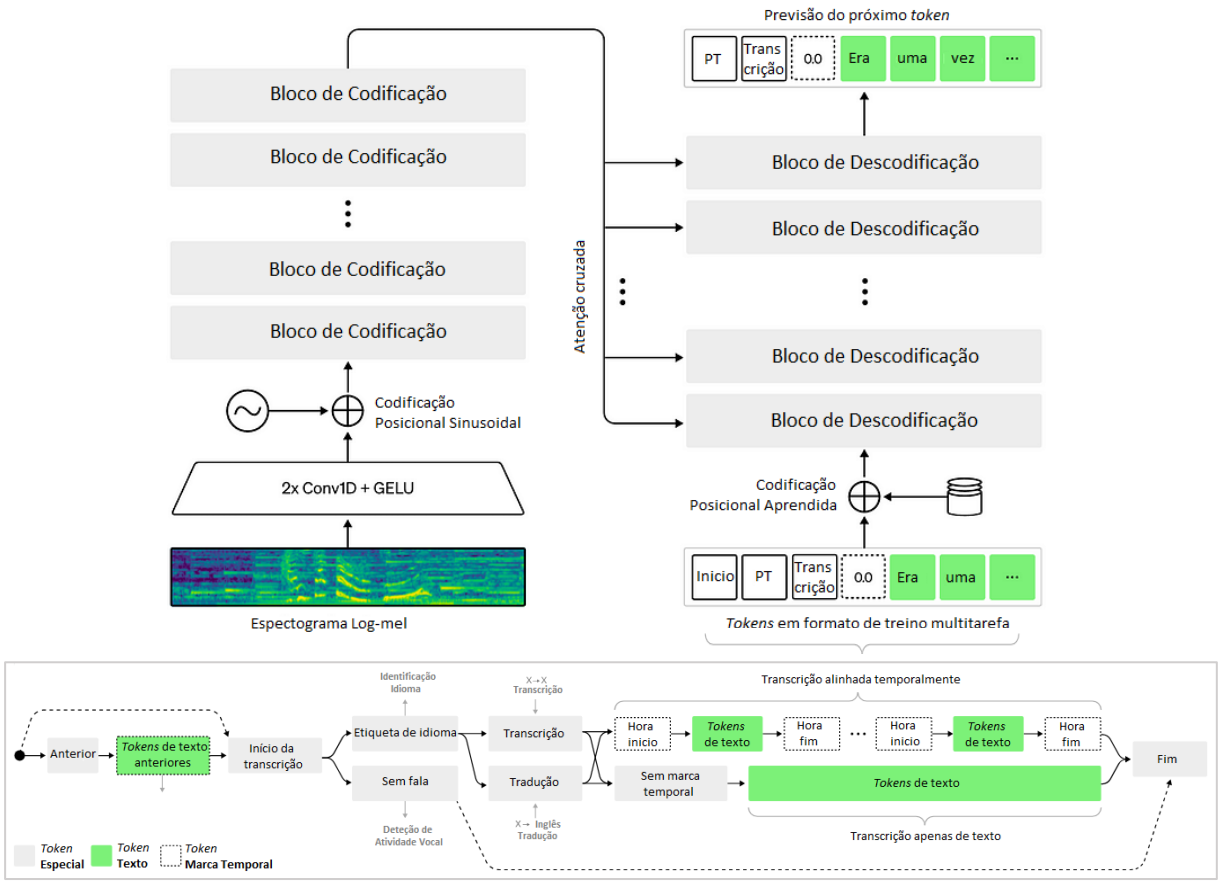


Figura 3 - Diagrama de funcionamento do Whisper [55], adaptado para português

A OpenAI afirma que, embora existam vários outros modelos que foram treinados com o objetivo de obter um bom desempenho em conjuntos de dados ou conjuntos de testes específicos, com o Whisper mudaram o foco para o pré-treino de uma forma supervisionada, utilizando conjuntos de dados maiores criados através da combinação de dados extraídos e filtrados através de grandes conjuntos de dados compilados e filtrando os dados utilizáveis através de heurísticas inteligentes. Estes modelos tendem a ser mais robustos e a generalizar muito mais eficazmente para casos de utilização no mundo real. Embora se possa verificar que alguns modelos superam o Whisper em conjuntos de testes individuais, ao escalar o seu pré-treino pouco supervisionado, a equipa da OpenAI conseguiu obter resultados muito impressionantes sem a utilização de técnicas de auto-supervisão e auto-treino habitualmente utilizadas pelos atuais modelos ASR de última geração.

Cerca de um terço do conjunto de dados utilizado para o treino não estava em inglês. O treino do Whisper envolveu a alternância entre duas tarefas, a transcrição do áudio na sua língua original e sua tradução para inglês. A equipa da OpenAI considerou que este estilo de treino era uma técnica eficaz para o Whisper aprender a tradução de voz para texto, o que resultou num desempenho superior aos métodos de treino supervisionado utilizados pelos modelos atuais mais avançados, quando testados no corpus multilingue CoVoST2 [56] para tradução em inglês.

Outra grande vantagem do Whisper é a ausência de subscrições ou licenças para o seu uso. Isto significa que indivíduos, empresas, organizações e investigadores independentes podem aproveitar as capacidades do Whisper sem terem de fazer grandes investimentos em licenças ou produtos caros.

É possível executar o Whisper localmente, ou seja, há um maior controlo sobre os dados que estão a ser processados, o que é particularmente relevante em cenários em que a privacidade dos dados é uma preocupação fundamental. Os dados de áudio não precisam de ser enviados para servidores externos para serem processados, o que se revela como vantagem significativa em termos de privacidade e segurança. Apesar da maioria das intervenções da AR serem públicas, algumas são reservadas pelo que ter a possibilidade de realizar a transcrição localmente é uma vantagem importante. Acresce ainda que o facto de poder ser executada localmente, permite uma maior personalização e adaptação do modelo para satisfazer as necessidades particulares da AR.

O desempenho do Whisper varia muito consoante a língua. A Figura 4 mostra uma análise de WER (taxa de erro de palavras) por idiomas do conjunto de dados Fleurs ao usar o modelo *large-v2*, onde se verifica que a língua portuguesa é das mais baixas, com 4,3% (quanto menor o WER, melhor o desempenho).

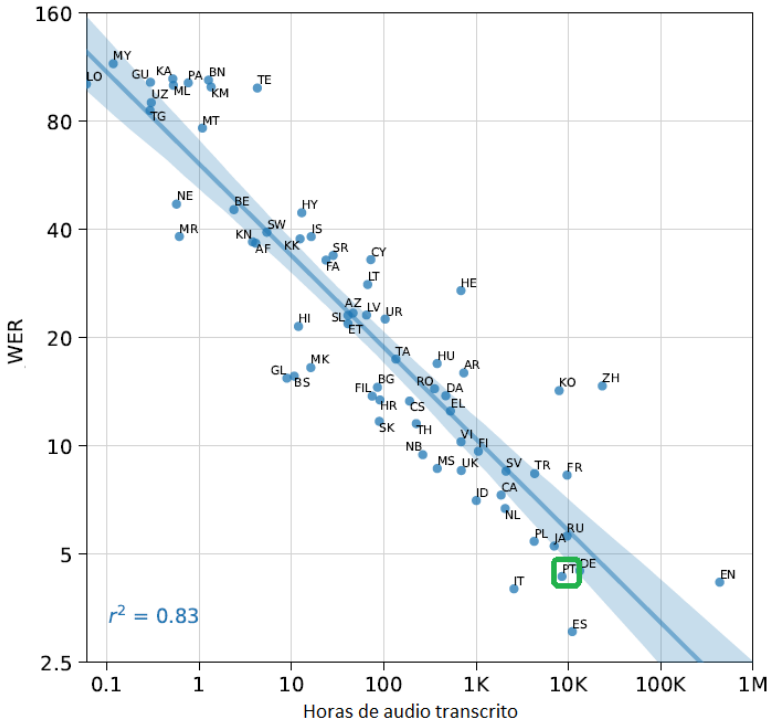


Figura 4 - Taxa de erro WER do modelo Whisper large-v2 ao usar o dataset Fleurs [55]

2.4.2. WhisperX

Os modelos de reconhecimento automático de fala em larga escala e com supervisão ligeira, como o Whisper, têm demonstrado resultados notáveis no reconhecimento de discurso em diversos domínios e idiomas. No entanto, os marcadores temporais (*timestamps*) associados a cada transcrição não só tendem a ser imprecisos como não estão disponíveis ao nível de cada palavra. Adicionalmente, a sua utilização em áudio extenso através de transcrição em *buffer* impede a inferência em lote devido à sua natureza sequencial. Para ultrapassar os desafios mencionados, Bain et al. [57] desenvolveram e disponibilizaram o WhisperX [58], um sistema de reconhecimento de voz com precisão temporal que fornece registos temporais ao nível das palavras, através da deteção de atividade vocal e alinhamento forçado de fonemas.

Assim, o WhisperX é um sistema desenhado para a transcrição eficiente da fala de áudios longos com alinhamento temporal ao nível das palavras. O áudio de entrada é primeiro segmentado com deteção de atividade de voz (VAD) e depois cortado e fundido em pedaços de entrada de aproximadamente 30 segundos com limites que se situam em regiões de fala minimamente ativas. Os segmentos resultantes são então transcritos em paralelo com o Whisper, e forçados a alinhar com um modelo de reconhecimento de fonemas para produzir registos de tempo precisos ao nível da palavra com um elevado rendimento.

O reconhecimento automático de fala (ASR) baseado em fonemas refere-se a uma abordagem de ASR que utiliza fonemas como unidades básicas de reconhecimento. Um fonema é a menor unidade sonora de uma palavra e é crucial para entender a pronúncia e o significado das palavras. Os sistemas ASR baseados em fonemas são treinados para reconhecer padrões fonéticos em áudio e mapeá-los para texto.

O WhisperX em particular utiliza o modelo Wav2Vec2.0, descrito por Baevski et al. [20], que é um modelo de aprendizagem profunda, desenvolvido pela Facebook AI Research, treinado para converter ondas sonoras em representações vetoriais que podem ser utilizadas para tarefas de ASR. Este modelo é conhecido pela sua eficácia em tarefas de ASR, e capaz de alcançar desempenho de ponta com menos dados de treino supervisionado, em comparação com abordagens tradicionais.

No final do processo o WhisperX procede ainda ao alinhamento de áudio, processo que mapeia os segmentos de áudio para as palavras ou fonemas correspondentes no texto, ação fundamental para utilizações como legendas automáticas, indexação de áudio, e outras aplicações que requerem uma sincronização precisa entre áudio e texto.

O WhisperX permite ainda a diarização, um processo de segmentação e identificação automática de diferentes oradores numa gravação de áudio. O objetivo da diarização do orador é dividir o fluxo de áudio em segmentos homogéneos, em que cada segmento corresponde a um orador específico ou a um turno de orador. Por outras palavras, o objetivo é responder à pergunta "Quem falou quando?" ao longo de uma gravação de áudio.

A diarização é feita com base em pyannote.audio, um conjunto de ferramentas de código aberto escrito em Python por Bredin et al. [59].

A Figura 5 representa um diagrama de funcionamento do WhisperX [57], adaptado para português, relativo ao alinhamento dos *timestamps* com a transcrição a ser realizada pelo Whisper.

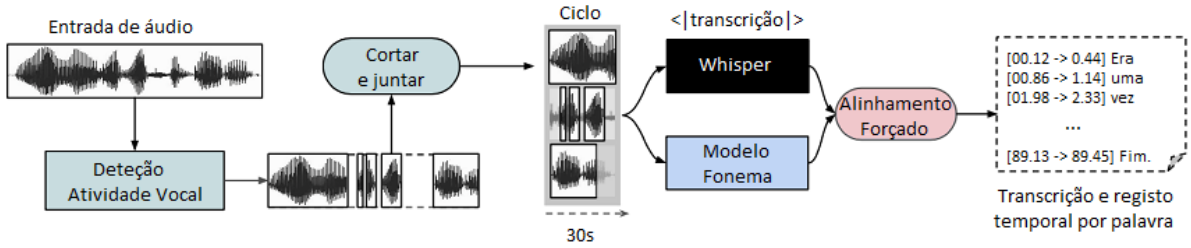


Figura 5 - Diagrama de funcionamento do WhisperX [57], adaptado para português

2.5. Conclusão

Ao analisar trabalho relacionado, constata-se que a criação de bases de dados específicas para tarefas de transformação de áudio em texto tem sido uma estratégia amplamente adotada na comunidade científica. Esta abordagem, embora eficaz em muitos casos, apresenta desafios significativos. A construção de tais *corpora* exige investimentos substanciais em termos de recursos humanos, financeiros e computacionais. Além disso, a necessidade de dados de fala de alta qualidade e em grande quantidade, muitas vezes escassos ou de difícil acesso, torna o processo ainda mais complexo e demorado.

A disponibilização de modelos de linguagem de grande dimensão (LLMs), como o Whisper, representa uma viragem paradigmática nesta área. Estes modelos, treinados em vastos conjuntos de dados e capazes de generalizar para uma ampla variedade de tarefas e idiomas, oferecem uma alternativa promissora à abordagem tradicional baseada em *corpora*.

Assim, apesar da criação de *corpus* específicos continuar a ter o seu lugar na pesquisa e desenvolvimento de sistemas ASR, a ascensão dos LLMs oferece uma via mais ágil e económica para a implementação de soluções de transcrição automática. Para a Assembleia da República e outras instituições semelhantes, esta evolução representa uma oportunidade valiosa para melhorar a acessibilidade e a eficiência dos seus registos parlamentares, beneficiando tanto a instituição como o público em geral.

CAPÍTULO 3.

Desenho e Desenvolvimento

O sistema de transcrição automática tem como objetivo principal a automatização de tarefas que atualmente a equipa de redatores da AR tem de realizar de forma manual. Como tal o Sistema de Transcrição Automática para a Assembleia da República (STAAR) foi desenvolvido para responder às necessidades específicas da transcrição de debates parlamentares na AR. Este capítulo é dedicado à descrição detalhada do STAAR, desde a identificação dos requisitos necessários, a definição da arquitetura, descrição do sistema e finalmente as tarefas relativas ao desenvolvimento efetuado.

3.1. Identificação dos requisitos

A presente secção descreve a fase inicial do processo de desenvolvimento do STAAR com foco nos debates parlamentares. É apresentada uma análise detalhada das necessidades e requisitos específicos que a solução deve atender, garantindo sua eficácia e adequação ao contexto da AR.

De acordo com a metodologia DSRM, referida em 1.4, a identificação dos requisitos é um passo fundamental na criação de um sistema. Entre dezembro de 2022 e fevereiro de 2023 foram realizadas cinco reuniões formais com os principais *stakeholders*, como elementos da equipa que produz o diário da AR, elementos da equipa que grava os áudios das sessões e chefias das áreas envolvidas no processo. Foram ainda realizados diversos contactos com a equipa de transcrição no sentido de conhecer em maior detalhe o funcionamento e forma de transcrição dos debates parlamentares.

Através destas reuniões e contactos, de experiências anteriores com outros *softwares* de transcrição, bem como do conhecimento do funcionamento dos sistemas de informação da AR, foi possível realizar uma avaliação abrangente das necessidades e desafios específicos do contexto parlamentar, que resultou numa lista detalhada de requisitos no âmbito do presente trabalho de projeto. Esta lista, apresentada na Tabela 4, serviu como alicerce fundamental para o desenvolvimento do STAAR, como forma de assegurar que o sistema fosse construído com uma base sólida de necessidades reais e expectativas claras.

Foram ainda considerados aspetos linguísticos, técnicos e operacionais, incluindo a velocidade e precisão da transcrição, a capacidade de processar áudios já existentes, a utilização dos recursos já existentes, bem como a flexibilidade de expansão para transcrição de outras atividades parlamentares, como reuniões de comissões parlamentares. Outro requisito considerado importante foi a capacidade de lidar com desafios acústicos específicos, como ruído de fundo, palmas ou outros oradores a falar em simultâneo.

Tabela 4 - Problemas e requisitos identificados

#	Problema	Requisito
1	Redatores estão muito tempo ocupados com a transcrição manual	A solução deve transcrever automaticamente a fala em texto, eliminando a necessidade de digitação manual
2	Há diferenças entre o que é dito e o que é escrito, para ser legível	A transcrição automática deve produzir um texto que não seja uma transcrição *verbatim*, mas sim um registo legível e compreensível das intervenções parlamentares
3	Há muitas macros desenvolvidas para tratamento de texto	A solução deve permitir a edição e formatação do texto transcrito diretamente no Word, facilitando a revisão e ajustes necessários
4	No caso das sessões plenárias a transcrição deve estar disponível de forma rápida, após o áudio ter sido disponibilizado	A solução deve processar os áudios o mais rapidamente por forma a disponibilizá-los, no máximo, até 10 minutos depois de estarem disponíveis
5	Pretende-se que os utilizadores recorram aos equipamentos que já têm para tratar o texto transcrito	A solução deve ser compatível com os equipamentos já existentes no ambiente parlamentar, evitando a necessidade de investimento em novos equipamentos para os utilizadores
6	O áudio deve ser controlado por um *hardware* próprio (pedaleira)	A solução deve ser capaz de suportar o uso de uma pedaleira como um *hardware* específico para controle de pausas e outras ações durante a transcrição
7	Geralmente cada áudio contém diversos oradores	A solução deve ser capaz de identificar sempre que há mudança de orador e associar o texto transcrito a cada um
8	Os áudios a transcrever encontram-se em diversas fontes	A solução deve ter a possibilidade de obter os áudios de várias localizações
9	Para além dos áudios automáticos pretende-se transcrever outros áudios *ad hoc*	A solução deve ser capaz de transcrever também áudios introduzidos de forma manual pelos utilizadores
10	Não existem modelos de linguagem públicos em português europeu	A solução deve ser desenvolvida com atenção às particularidades da língua portuguesa, levando em consideração a falta de um *corpus* amplo e específico para o reconhecimento automático de fala nesse idioma

11	Antes de adquirir ou implementar qualquer solução pretende-se testar as suas capacidades e avaliar se é exequível	A solução deve permitir a realização de avaliações e testes para garantir sua eficácia e adequação antes da implementação completa
12	Numa fase inicial pretende-se transcrever sessões plenárias, mas deve ser prevista a transcrição de reuniões de comissão	A solução deve ser projetada com a capacidade de expansão para transcrição de reuniões de comissões parlamentares, além dos debates no plenário
13	Alguns termos usados devem ser automaticamente substituídos por outros, podendo os utilizadores alterar esses termos	A solução deve permitir a criação e personalização de dicionários de substituição, facilitando a correção automática de termos específicos e jargões parlamentares
14	A transcrição deve ser o mais exata possível	A solução deve visar um baixo Word Error Rate (WER), no máximo 15%, garantindo a precisão e a confiabilidade da transcrição automática
15	Alguns áudios são confidenciais ou reservados, pelo que a transcrição deve ser feita usando recursos da AR sem enviar para soluções em *cloud*	A solução deve processar os áudios na infraestrutura da AR, garantindo a segurança e a confidencialidade dos dados durante o processo de transcrição
16	A solução deve estar disponível em teletrabalho	A solução deve oferecer a capacidade de trabalho remoto, permitindo que os transcritores acedam e utilizem o sistema de qualquer localização geográfica, desde que tenham as permissões adequadas
17	Existem vários níveis de acesso às transcrições	A solução deve fornecer diferentes níveis de permissões de acesso, garantindo que apenas utilizadores autorizados tenham acesso aos recursos e funcionalidades específicas do sistema, como forma de manter a segurança e a confidencialidade dos dados
18	Muitas vezes os áudios têm barulho de fundo, seja por palmas, seja por interjeições de outros deputados	A solução deve ser capaz de lidar com ruídos de fundo durante a transcrição, como palmas, outras pessoas ou qualquer outro tipo de interferência sonora. É importante que o sistema seja capaz de filtrar e distinguir claramente a fala relevante, garantindo a precisão da transcrição em ambientes acústicos desafiadores

3.2. Desenho da arquitetura

Para dar resposta às necessidades identificadas, foi desenhada uma arquitetura modular, pensada para otimizar o processo de transcrição de intervenções parlamentares. Esta arquitetura foi estruturada em quatro etapas sequenciais: Recolha de áudios, Processamento de áudio, Tratamento de texto e Armazenamento das transcrições, tal como ilustrado na Figura 6.

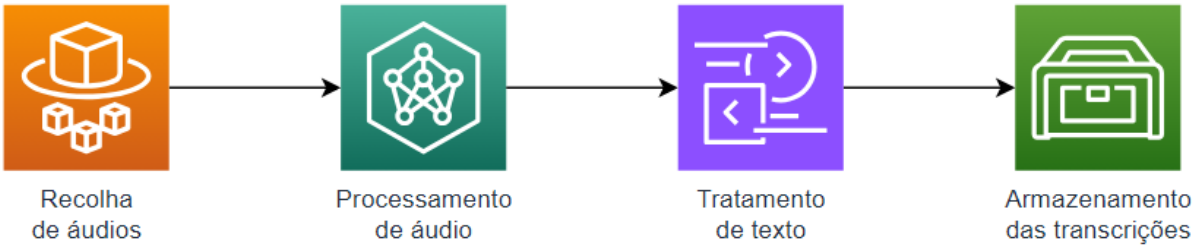


Figura 6 - Arquitetura conceptual do STAAR

Cada etapa foi desenhada para dar resposta a um conjunto específico de necessidades e desafios, garantindo uma transição fluida e eficiente do áudio para texto:

- Recolha de áudios - Esta etapa serve como ponto de partida, onde devem ser identificados e recolhidos áudios de diversas fontes parlamentares (ex: sessões plenárias, reuniões de comissão, etc.);
- Processamento de áudio - Esta etapa representa a essência do sistema, cujo principal objetivo é transformar as intervenções parlamentares, registadas em formato de áudio, em transcrições textuais precisas e compreensíveis;
- Tratamento de texto - Uma vez transcritos, os textos devem passar por um refinamento, para assegurar que as transcrições são claras, precisas e alinhadas com as normas e convenções parlamentares;
- Armazenamento das transcrições - As transcrições devem ser armazenadas de forma organizada, garantindo fácil acesso e gestão. Para além de repositório, esta etapa deve ainda manter um registo detalhado de todas as transcrições, por forma a fornecer uma base robusta para análises futuras e gestão de recursos.

O desenho da arquitetura do STAAR foi cuidadosamente planeado para garantir um processo de transcrição eficiente e preciso, alinhado com as necessidades e especificidades do ambiente parlamentar. A modularidade e a sequência lógica das etapas asseguram que o sistema é escalável, robusto e adaptável a futuras necessidades e desafios.

3.3. Desenvolvimento do sistema

O desenvolvimento do Sistema de Transcrição Automática para a Assembleia da República (STAAR) é uma tarefa meticulosa que visa transformar a maneira como os debates parlamentares são transcritos e documentados. Nesta fase, as ideias concebidas nas etapas anteriores são materializadas através de código, algoritmos e interfaces de utilizador intuitivas. Esta secção detalha o percurso de desenvolvimento do STAAR e as decisões tomadas para cumprir os objetivos estabelecidos.

A Figura 7 detalha as etapas e módulos envolvidos no processo de transcrição de áudios parlamentares, o que permite uma visualização clara da sequência de operações, desde a captação dos áudios até à criação de um repositório organizado de transcrições finalizadas.

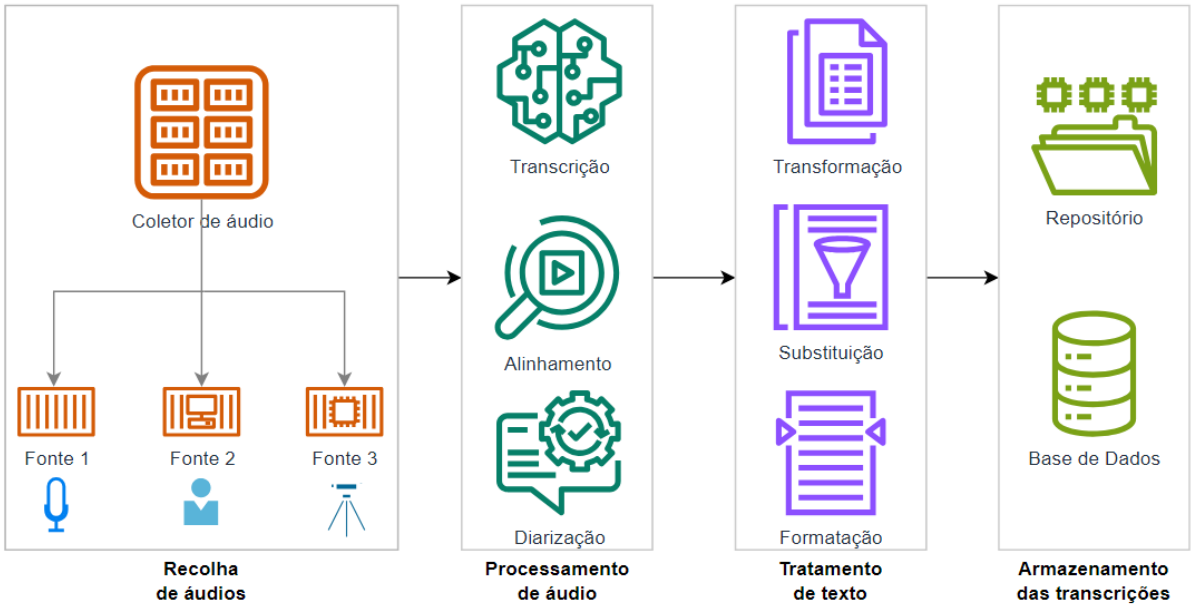


Figura 7 - Arquitetura detalhada do STAAR

Importa salientar que as diversas etapas e módulos do STAAR foram desenvolvidos em momentos distintos do projeto. Este desenvolvimento faseado permitiu uma implementação e revisão contínua por parte dos utilizadores, para assegurar que o sistema evoluía de acordo com as necessidades identificadas, tal como detalhado na secção 4.1 sobre as fases de implementação.

3.3.1. Recolha de áudios

Esta etapa representa o ponto de partida do processo, responsável por reunir áudios provenientes de diferentes fontes. Abrangendo sessões plenárias, comissões parlamentares e outros eventos relevantes, a Recolha de áudios estabelece as bases para a conversão subsequente do áudio para

texto. Nesta etapa é necessário ligar a várias fontes de áudio, determinar se os áudios já possuem transcrição e, caso não tenham, disponibilizá-los para a etapa seguinte, o Processamento de áudio.

O módulo desenhado para cumprir este objetivo, Coletor de áudios, foi materializado num script desenvolvido com a linguagem de programação Python [60], conhecida pela sua eficiência e versatilidade, especialmente em tarefas de manipulação de dados. A Figura 8 representa o processo da Recolha de Áudios, ilustrado através de um esquema BPMN [61], para uma visualização clara e estruturada, que é posteriormente descrito em maior pormenor.

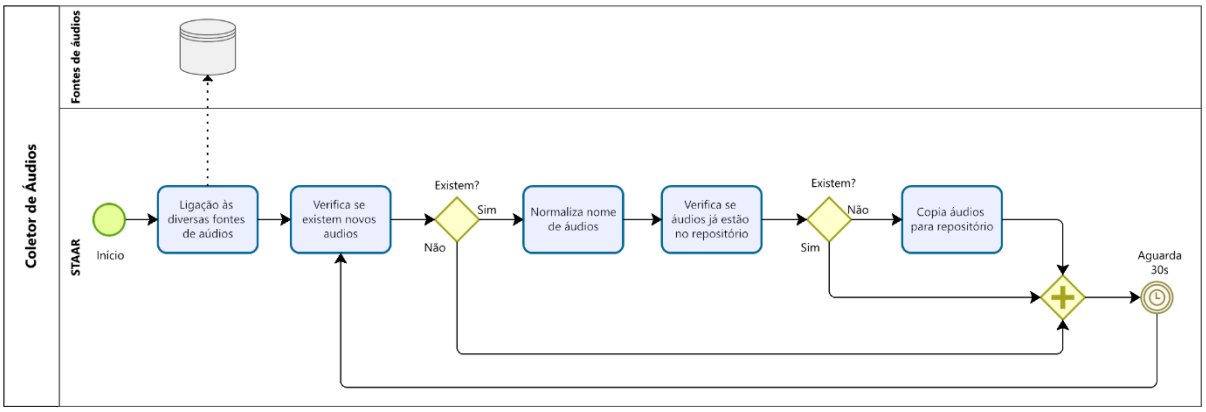


Figura 8 - Esquema BPMN do Coletor de Áudios

O processo de recolha de áudios inicia-se com a ligação às diversas fontes onde os áudios estão armazenados, por exemplo através de partilhas de rede. Estes áudios estão organizados em pastas, com os nomes representando o órgão correspondente, seja plenário, comissão, evento ou outros. É importante destacar que, devido à organização do trabalho parlamentar, os áudios de plenário são segmentados em fragmentos de 15 minutos, onde uma sessão plenária típica tem a duração de cerca de 3 horas. Assim, a cada quarto de hora, o sistema de gravação gera um novo ficheiro de áudio que é colocado na fonte de áudio. Já os áudios de comissões parlamentares e eventos podem ser segmentados de forma idêntica ou existir como um único ficheiro de áudio.

Após estabelecer a ligação com as fontes, o sistema identifica os ficheiros modificados nos últimos 30 dias, um prazo estendido para acomodar possíveis atrasos na disponibilização dos áudios. Os áudios identificados são então submetidos a um processo de normalização, que visa homogeneizar os nomes dos ficheiros, facilitando a gestão e futura referenciação dos mesmos. Este processo utiliza a data e hora de modificação do ficheiro (formato yyyy-dd-mm_hh-mm-ss) para garantir uma nomenclatura consistente e evitar duplicações.

Posteriormente, o sistema verifica no repositório de textos transcritos se o áudio já foi processado, através da comparação dos nomes dos ficheiros. Se uma transcrição correspondente for identificada, o áudio é descartado. Caso contrário, é transferido para um repositório central, onde fica a aguardar

a transcrição. Esta verificação evita duplicações e garante que apenas áudios não transcritos são processados.

Por último, este ciclo de verificação e coleta é executado a cada 30 segundos para assegurar que os áudios são recolhidos e disponibilizados para o processo de transcrição de forma quase imediata, atendendo à necessidade de transcrições rápidas no ambiente parlamentar (requisito #4).

3.3.2. Processamento de áudio

A etapa de Processamento de áudio é o epicentro onde a conversão de áudio para texto é realizada de forma a transformar o discurso falado nas intervenções parlamentares em transcrições textuais coesas e bastante precisas. O processamento do áudio é realizado em três fases distintas: Transcrição, Alinhamento e Diarização. Cada uma dessas fases tem um papel essencial para garantir a precisão e a utilidade da transcrição final, cujo funcionamento é representado através de um esquema BPMN [61] na Figura 9, o que permite obter uma visão sistematizada do seu funcionamento.

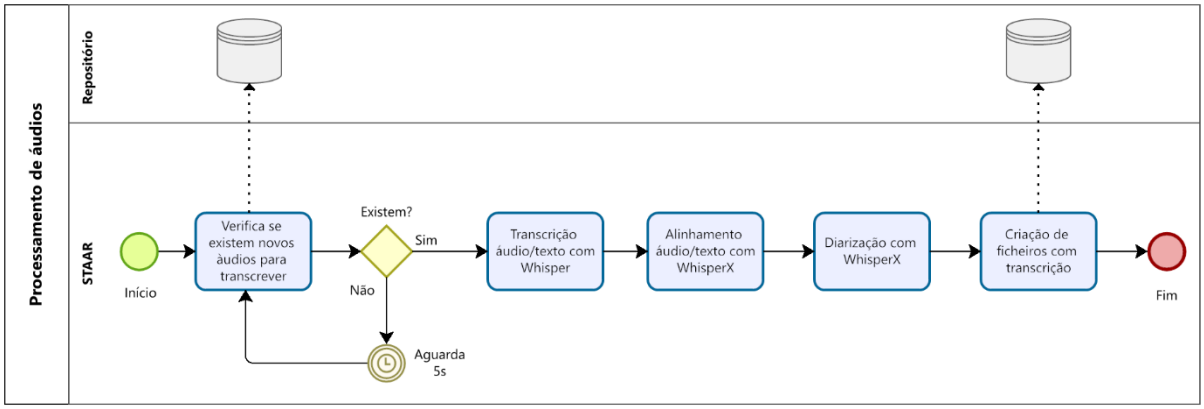


Figura 9 - Esquema BPMN do Processamento de Áudios

3.3.2.1. Transcrição

Este módulo é o núcleo central do Sistema de Transcrição Automática (STAAR). A principal função é transformar as intervenções parlamentares, registadas em formato de áudio, em transcrições textuais fidedignas e como tal, a maior complexidade deste módulo é a compreensão e interpretação precisa do discurso oral.

Para realizar esta tarefa, e fruto da pesquisa realizada no Capítulo 2, optou-se por aplicar modelos de linguagem baseados em Inteligência Artificial (IA), que afirmam permitir uma transcrição fluida do discurso falado para o formato escrito. Através de um modelo de linguagem de grande dimensão (LLM, do inglês *Large Language Model*), treinado com uma vasta gama de dados linguísticos, estes modelos

são capazes de reconhecer palavras, frases e entonações específicas. Ao analisar os padrões sonoros e contextuais, o LLM traduz o áudio em texto, criando uma representação textual do discurso original.

O modelo escolhido para esta função foi o sistema de Reconhecimento Automático de Fala (ASR) denominado Whisper, desenvolvido pela OpenAI [50]. Entre os diversos modelos disponibilizados pelo Whisper, optou-se pelo uso do modelo *large-v2*, devido à sua capacidade de proporcionar transcrições com a menor taxa de erro por palavra.

Foi desenvolvida uma programação em python para verificar a existência de áudios no repositório central temporário. Quando identificado um áudio, este é submetido ao Whisper, que, através da análise dos primeiros 30 segundos do áudio, deteta o idioma do áudio e realiza a transição com base nessa língua. Esta funcionalidade é grande utilidade, visto que, apesar da predominância do idioma português, há ocasiões em que outros idiomas são utilizados, tornando a capacidade multilíngue do Whisper um recurso valioso.

Ao longo do processo são recolhidas informações, como data/hora de início e fim do processo da transcrição que mais tarde serão colocados em base de dados, para efeitos de avaliação de performance do sistema.

As Figura 26 e a Figura 27 apresentam extratos, respetivamente, do código desenvolvido para executar a função de procura de ficheiros por transcrever e a transcrição propriamente dita dos áudios.

3.3.2.2. Alinhamento

Este módulo é responsável por sincronizar palavra por palavra com o áudio original, ao estabelecer marcas temporais (*timestamps*) precisas para cada palavra transcrita. Este alinhamento rigoroso assegura que o texto e o áudio permaneçam em sintonia, preservando a sequência temporal exata do áudio transcrito.

Embora o modelo LLM Whisper crie transcrições com *timestamps*, o alinhamento entre som/tempos não é rigoroso. Esta precisão no alinhamento é importante para manter a sequência temporal exata do áudio transcrito. Assim, face às referidas lacunas do Whisper no que diz respeito ao alinhamento do som com *timestamps* foi necessário encontrar soluções para o problema. Uma solução encontrada foi a utilização do WhisperX [57], uma extensão do Whisper, disponível no GitHub [58], que apresenta várias vantagens, tornando-o numa escolha válida para as fases de alinhamento, representado na Figura 10, e diarização.

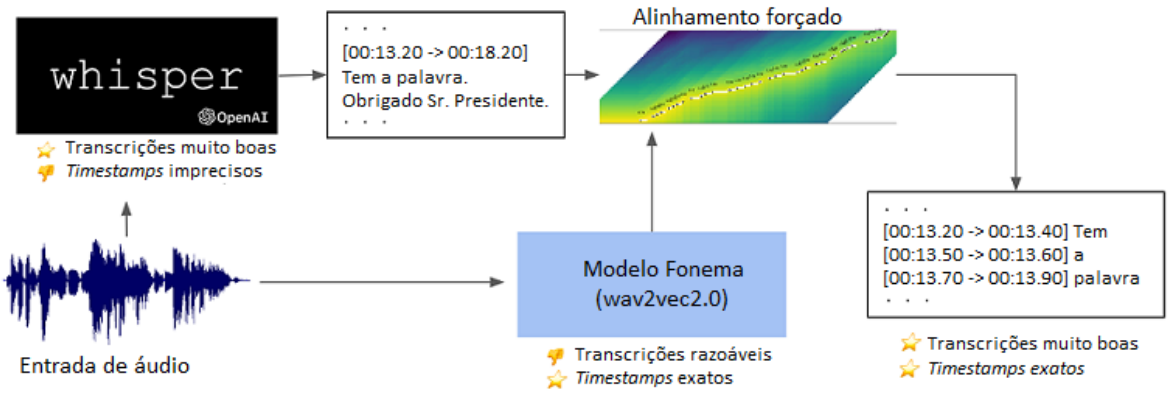


Figura 10 - Vantagens de utilizar WhisperX [57], adaptado para português

O WhisperX não só dá resposta aos problemas referidos, como consegue realizar as transições em bastante menos tempo, mesmo com modelo "large-v2" do Whisper, tem menor requisitos de computacionais, reduz os erros relacionados com alucinações durante o processo de deteção de atividade de voz (VAD), como ainda tem funções de reconhecimento do orador, diarização, que o Whisper não disponibiliza.

O processo de alinhamento é executado através uma função desenvolvida em código python (Figura 28), criado a partir da documentação do produto. Também neste processo são recolhidos dados para calcular a performance associada ao alinhamento.

3.3.2.3. Diarização

Por último o módulo de Diarização distingue e numera as vozes presentes no áudio, atribuindo identificadores aos diferentes oradores. Apesar de não identificar o nome exato do orador, a contabilização do número de oradores em muito auxilia o trabalho de revisão da transcrição, uma vez que por vezes os áudios/textos são bastante longos pelo que um contexto por orador é uma ajuda importante.

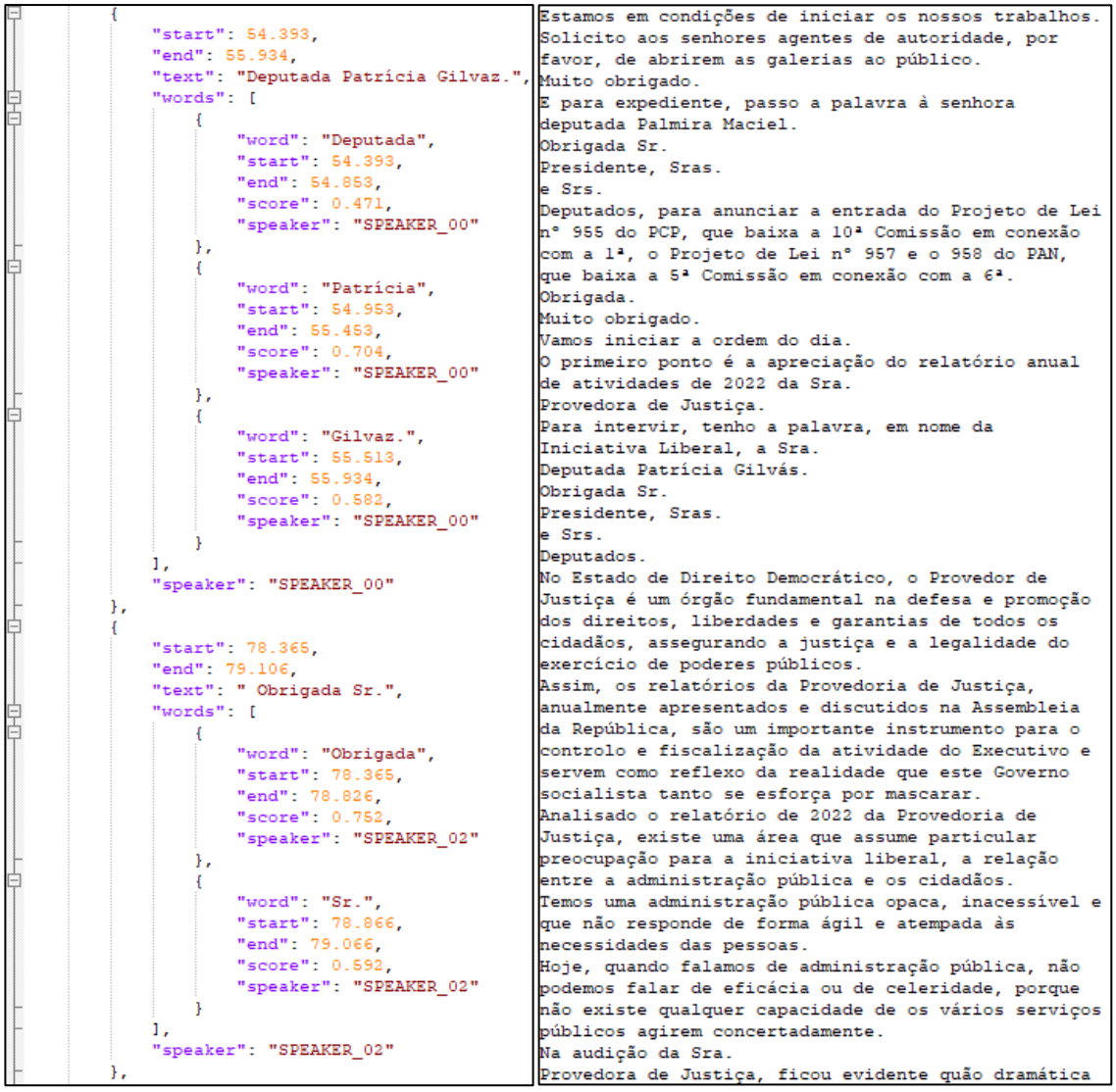
Para a diarização é usado novamente o WhisperX que, ao disponibilizar os *timestamps* exatos torna possível o processo de atribuição de oradores a cada segmento de texto encontrado.

O processo de diarização é uma função do código python desenvolvido (Figura 29), criado a partir da documentação do WhisperX e onde são recolhidos os dados finais para calcular a *performance* associada a esta etapa.

O WhisperX permite gravar o resultado das fases de transcrição, alinhamento e diarização em diferentes formatos, nomeadamente “json”, “srt”, “tsv”, “txt” e “vtt”, que podem ser usados por várias plataformas, como por exemplo para transcrição ou legendagem.

Para o efeito da transcrição dos debates parlamentares, no final do Processamento de áudio apenas são guardados os ficheiros “json” e “srt”, exemplificados na Figura 11. O formato “json” é o

mais completo na medida em que contém a transcrição com *timestamps* ao nível de cada palavra de forma estruturada, enquanto o "srt" apenas tem *timestamps* ao nível das frases, sendo este o formato utilizado na fase seguinte do processo de transcrição.




```
        {
            "start": 54.393,
            "end": 55.934,
            "text": "Deputada Patrícia Gilvaz.",
            "words": [
                {
                    "word": "Deputada",
                    "start": 54.393,
                    "end": 54.853,
                    "score": 0.471,
                    "speaker": "SPEAKER_00"
                },
                {
                    "word": "Patrícia",
                    "start": 54.953,
                    "end": 55.453,
                    "score": 0.704,
                    "speaker": "SPEAKER_00"
                },
                {
                    "word": "Gilvaz.",
                    "start": 55.513,
                    "end": 55.934,
                    "score": 0.582,
                    "speaker": "SPEAKER_00"
                }
            ],
            "speaker": "SPEAKER_00"
        },
        {
            "start": 78.365,
            "end": 79.106,
            "text": " Obrigada Sr.",
            "words": [
                {
                    "word": "Obrigada",
                    "start": 78.365,
                    "end": 78.826,
                    "score": 0.752,
                    "speaker": "SPEAKER_02"
                },
                {
                    "word": "Sr.",
                    "start": 78.866,
                    "end": 79.066,
                    "score": 0.592,
                    "speaker": "SPEAKER_02"
                }
            ],
            "speaker": "SPEAKER_02"
        },
```


```
Estamos em condições de iniciar os nossos trabalhos.
Solicito aos senhores agentes de autoridade, por
favor, de abrirem as galerias ao público.
Muito obrigado.
E para expediente, passo a palavra à senhora
deputada Palmira Maciel.
Obrigada Sr.
Presidente, Sras.
e Srs.
Deputados, para anunciar a entrada do Projeto de Lei
nº 955 do PCP, que baixa a 10ª Comissão em conexão
com a 1ª, o Projeto de Lei nº 957 e o 958 do PAN,
que baixa a 5ª Comissão em conexão com a 6ª.
Obrigada.
Muito obrigado.
Vamos iniciar a ordem do dia.
O primeiro ponto é a apreciação do relatório anual
de atividades de 2022 da Sra.
Provedora de Justiça.
Para intervir, tenho a palavra, em nome da
Iniciativa Liberal, a Sra.
Deputada Patrícia Gilvás.
Obrigada Sr.
Presidente, Sras.
e Srs.
Deputados.
No Estado de Direito Democrático, o Provedor de
Justiça é um órgão fundamental na defesa e promoção
dos direitos, liberdades e garantias de todos os
cidadãos, assegurando a justiça e a legalidade do
exercício de poderes públicos.
Assim, os relatórios da Provedoria de Justiça,
anualmente apresentados e discutidos na Assembleia
da República, são um importante instrumento para o
controlo e fiscalização da atividade do Executivo e
servem como reflexo da realidade que este Governo
socialista tanto se esforça por mascarar.
Analisado o relatório de 2022 da Provedoria de
Justiça, existe uma área que assume particular
preocupação para a iniciativa liberal, a relação
entre a administração pública e os cidadãos.
Temos uma administração pública opaca, inacessível e
que não responde de forma ágil e atempada às
necessidades das pessoas.
Hoje, quando falamos de administração pública, não
podemos falar de eficácia ou de celeridade, porque
não existe qualquer capacidade de os vários serviços
públicos agirem concertadamente.
Na audição da Sra.
Provedora de Justiça, ficou evidente quão dramática
```

Figura 11 - Exemplo de transcrição, após o Processamento de áudio, em formato "json" e "txt"

3.3.3. Tratamento de texto

A etapa de tratamento consiste em transformar as transcrições em bruto, geradas pelo sistema na etapa de Processamento de áudio, em documentos estruturados, legíveis e prontos para revisão. Este processo envolve várias atividades e técnicas que asseguram a clareza e a precisão do texto final, de forma a facilitar a sua revisão e utilização subsequente. Neste âmbito trata-se de transformar o resultado do Processamento de áudio (Figura 12), através de diversos processos detalhados nesta secção, num documento muito próximo daquele que a equipa de transcrição da AR produz (Figura 13).

```
1
00:00:00,049 --> 00:00:03,973
[SPEAKER_00]: Estamos em condições de iniciar os nossos trabalhos.

2
00:00:03,973 --> 00:00:09,017
[SPEAKER_00]: Solicito aos senhores agentes de autoridade, por favor, de abrirem as galerias ao
público.

3
00:00:09,017 --> 00:00:10,859
[SPEAKER_00]: Muito obrigado.

4
00:00:10,859 --> 00:00:14,803
[SPEAKER_00]: E para expediente, passo a palavra à senhora deputada Palmira Maciel.

5
00:00:15,964 --> 00:00:16,764
[SPEAKER_04]: Obrigada Sr.
```

Figura 12 - Exemplo de transcrição, antes do Processamento de texto, em formato "srt"

ORADOR_00: — Estamos em condições de iniciar os nossos trabalhos. Solicito aos senhores agentes de autoridade, por favor, de abrirem as galerias ao público. Muito obrigado. E para expediente, passo a palavra à Sr.ª Deputada Palmira Maciel.

ORADOR_04: — Obrigada, Sr. Presidente. Sras. e Srs. Deputados, para anunciar a entrada do Projeto de Lei n.º 955 do PCP que baixa a 10.ª Comissão em conexão com a 1.ª, o Projeto de Lei n.º 957 e o 958 do PAN, que baixa a 5.ª Comissão em conexão com a 6.ª. Obrigada.

ORADOR_00: — Muito obrigada. Vamos iniciar a ordem do dia. O primeiro ponto é a apreciação do relatório anual de atividades de 2022 da Sra. Provedora de Justiça. Para intervir, tem a palavra, em nome da Iniciativa Liberal, a Sra. Deputada Patrícia Gilvaz.

ORADOR_02: — Obrigada Sr. Presidente, Sras. e Srs. Deputados. No Estado de Direito Democrático, o Provedor de Justiça é um órgão fundamental na defesa e promoção dos direitos…

Figura 13 - Exemplo de transcrição, após o Tratamento de texto, em formato "docx"

Conforme se pode constatar na Figura 12, o texto que é obtido na etapa de Processamento de áudio é de difícil leitura uma vez que contém bastante informação desnecessária para quem tem de usar a transcrição para a criação do Diário da Assembleia da República (DAR). Após a etapa de Tratamento, o texto torna-se bastante mais legível, num formato muito semelhante ao DAR e pronto a ser revisto. Como se pode verificar na Figura 13, o texto é agrupado por orador e são substituídos

termos da gíria parlamentar (identificados a azul) entre outras alterações que são detalhadas nesta secção.

A Figura 14 representa, através de um esquema BPMN [61], a sequência das atividades contidas na etapa de Tratamento de texto.

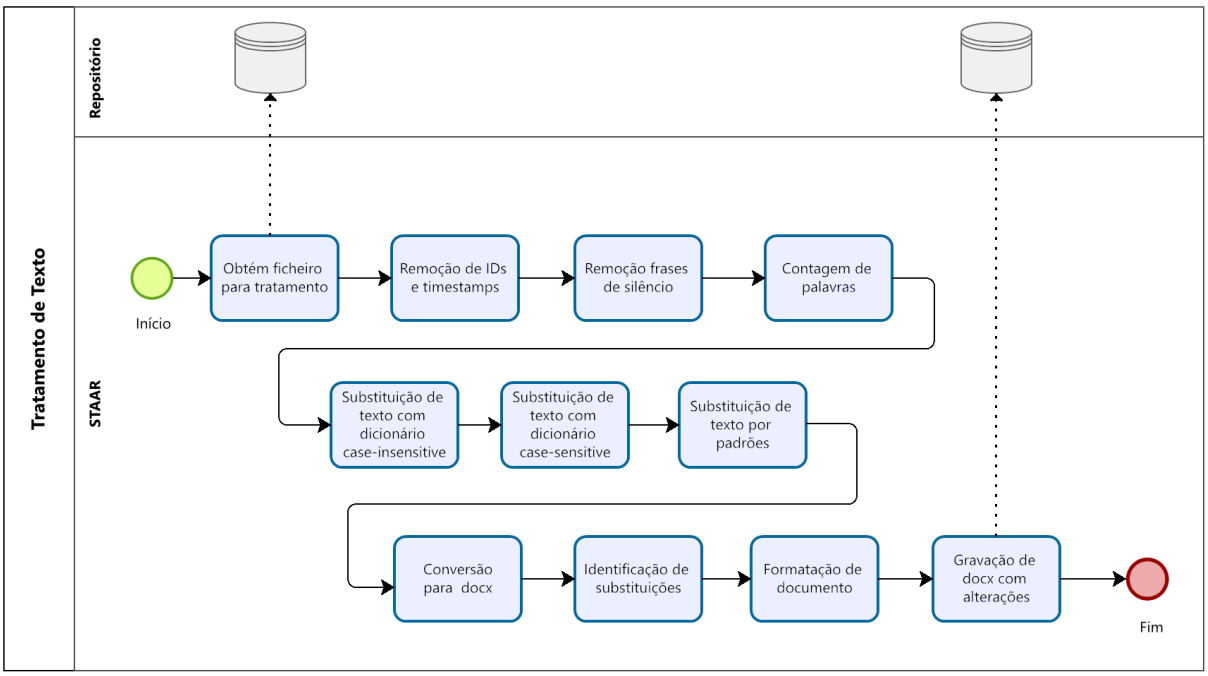


Figura 14 - Esquema BPMN do Tratamento de texto

3.3.3.1. Transformação

O propósito do módulo de Transformação é converter o texto em bruto, proveniente da etapa de Processamento de áudio, para um formato adequado. Esta etapa elimina informações desnecessárias, IDs e marcas temporais, frases aleatórias que surgem quando o áudio tem pausas muito longas, e foca-se em extrair apenas o conteúdo relevante para a elaboração do DAR. No final deste processo, é realizada uma contagem das palavras transcritas para fornecer dados estatísticos valiosos para análises futuras. Estas tarefas são realizadas através de código python desenvolvido para o efeito (Figura 30, Figura 31 e Figura 32).

3.3.3.2. Substituição

A função principal deste módulo é corrigir e ajustar palavras ou expressões que foram mal transcritas pelo sistema. Dada a singularidade da linguagem parlamentar, é comum que certos termos ou jargões sejam mal interpretados por sistemas automáticos. Este módulo incorpora regras linguísticas e um dicionário específico, alinhados às normas parlamentares, para determinar quais termos necessitam

de substituição e qual o grau de capitalização adequado, de acordo com as convenções linguísticas do ambiente parlamentar.

A atividade de substituição de texto é realizada através de um léxico, ou dicionário, composto por dois ficheiros de texto distintos. Um dos ficheiros é dedicado a substituições sem distinção entre maiúsculas e minúsculas (*case insensitive*), enquanto o outro engloba substituições sensíveis a maiúsculas e minúsculas (*case sensitive*). Esta distinção assegura uma correspondência mais precisa e contextual com os termos no texto transcrito.

Este dicionário, exemplificado na Tabela 5, contém um conjunto de palavras e expressões que são substituídas para melhorar a legibilidade e a clareza do texto, mantendo a integridade das ideias expressas durante as intervenções parlamentares.

Tabela 5 - Exemplo de termos a substituir de forma automática

Procurar	**Substituir por**
administração pública	Administração Pública
décima primeira comissão	11.ª Comissão
22.º governo	XXII Governo
senhor primeiro-Ministro	Sr. Primeiro-Ministro
hemiciclo	Hemiciclo

Os ficheiros de substituição são o reflexo acumulado da experiência da equipa de transcrição do Parlamento, contendo termos e expressões mais comuns que necessitaram de correção manual. Muitos destes termos pertencem à gíria ou ao léxico específico do ambiente parlamentar, o que os torna candidatos ideais para substituição automatizada.

Cada linha nos ficheiros de substituição contém um par de termos, o termo a ser procurado e o termo substituído, separados por uma vírgula. A lógica de substituição é executada por um script em Python (Figura 33), que percorre o texto transcrito. Ao identificar ocorrências dos termos especificados nos ficheiros de substituição, o script procede à substituição automática, refinando assim o texto. Quando uma substituição é efetuada, a palavra ou expressão substituída é prefixada e sufixada com "***" para que, mais tarde no processo de Tratamento de texto, se identifique quais as palavras substituídas.

Este processo não só melhora a precisão das transcrições, mas também economiza o tempo valioso da equipa de transcrição, minimizando a necessidade de intervenção manual em erros recorrentes.

3.3.3.3. Formatação

Este módulo é responsável pela formatação dos ficheiros de transcrição, a serem usados pelos revisores parlamentares, no formato AR que é necessário, com a identificação clara das alterações realizadas no módulo anterior e com um nome/localização que seja reconhecido, tanto pelos utilizadores como pelas aplicações que podem integrar com o STAAR, para se obter as transcrições.

Este módulo transforma os resultados brutos em documentos de qualidade prontos uso interno, de forma a permitir revisões, edições e formatação de acordo com as necessidades identificadas.

A primeira atividade nesta etapa é a conversão da transcrição para o formato docx. A decisão de adotar o formato docx para armazenar as transcrições resultou de uma análise estratégica baseada em vários requisitos identificados (Tabela 4). O formato docx é naturalmente compatível com o Microsoft Word, uma ferramenta amplamente adotada no ambiente parlamentar, permitindo que os utilizadores editem e formatem as transcrições diretamente, otimizando a eficiência do processo de revisão (requisito #3). Além disso, este formato assegura compatibilidade com os equipamentos já existentes no parlamento, evitando investimentos adicionais (requisito #5). Por último, a versatilidade do docx, suportado numa grande variedade de dispositivos, facilita o trabalho remoto, ao permitir a partilha e edição em tempo real por diversos utilizadores, independentemente da sua localização geográfica, desde que autorizados (requisito #16).

Do ponto de vista técnico esta conversão é feita através de uma função python (Figura 34) que utiliza a biblioteca python-docx.

Após a transcrição estar no formato docx, é iniciado o processo de Identificação de substituições, onde se identificada com a cor azul todos os termos substituídos na etapa de Substituição de texto. Esta coloração facilita a identificação, por parte do revisor parlamentar, das alterações resultantes da rotina automática de substituição. Assim, os revisores podem discernir com maior facilidade quais termos poderão necessitar de inclusão no dicionário de substituições, otimizando continuamente o processo. A função desenvolvida para executar esta tarefa pode ser consultada na Figura 35.

Por último, foi criada uma função (Figura 36) que alinha o texto para "justificar", para criar uma distribuição uniforme entre as margens, e altera o tipo de letra e no tamanho. Esta abordagem de formatação não só ajuda a equipa de transcrição a visualizar e trabalhar com o texto de forma mais eficaz, mas também serve como um passo preliminar importante na preparação do texto para a sua integração no DAR, garantindo que o documento final seja não só preciso, mas também esteticamente alinhado com os padrões parlamentares.

3.3.4. Armazenamento

O armazenamento é o repositório central onde as transcrições processadas são armazenadas, organizadas e geridas. A estrutura de armazenamento foi desenhada para garantir uma recuperação eficiente dos dados, bem como para facilitar o acesso e a gestão das transcrições por parte dos utilizadores e sistemas associados.

A Figura 15 representa, através de um esquema BPMN [41], a sequência das atividades contidas na etapa de armazenamento.

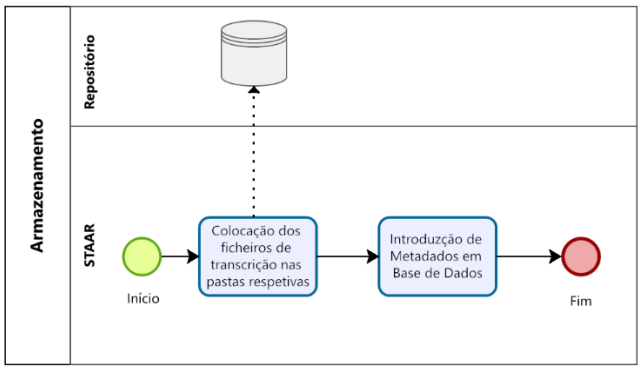


Figura 15 - Esquema BPMN do Armazenamento

Optou-se por um sistema de partilha de ficheiros em rede como a solução de armazenamento, dada a sua simplicidade e eficácia na facilitação do acesso e exploração das transcrições geradas pelo STAAR pelos utilizadores.

A utilização de partilha de rede como repositório oferece uma vantagem significativa em termos de acessibilidade e simplicidade para os utilizadores. Esta abordagem permite que os utilizadores possam aceder ao repositório de transcrições diretamente dos postos de trabalho que já possuem, sem a necessidade de instalar qualquer software adicional para facilitar o acesso. Esta estratégia minimiza as barreiras técnicas e operacionais e proporciona uma integração suave e eficiente com o ambiente de trabalho existente. Além disso, simplifica o processo de acesso às transcrições, ao tornar o STAAR um sistema mais amigável e fácil de adotar pelos utilizadores, enquanto reduz os requisitos de gestão e manutenção técnica associados à implementação de soluções de *software* adicionais para acesso ao repositório.

A estrutura do repositório é organizada de acordo com o tipo de áudio, sendo categorizada em plenário, comissões e eventos. Dentro de cada uma destas categorias, existem subpastas correspondentes a comissões ou eventos específicos, e subsequentemente subpastas organizadas por dia. Cada ficheiro transcrito é nomeado seguindo o formato “yyyy-dd-mm_hh-mm-ss” e é armazenado no formato docx, conforme detalhado anteriormente. Esta estrutura hierárquica, exemplificada na

Figura 16, facilita a localização e gestão das transcrições e proporciona uma organização lógica e intuitiva dos dados.

A utilização de partilhas de rede permite a implementação eficaz de controlos de acesso ao repositório. Dada a natureza potencialmente confidencial de algumas transcrições, é imperativo que o acesso seja rigorosamente controlado. As permissões de acesso são atribuídas de forma a garantir que apenas indivíduos autorizados possam aceder a transcrições específicas, protegendo assim a integridade e confidencialidade dos dados.

Foi considerado o desenvolvimento de uma API para permitir a obtenção de ficheiros de transcrição em diferentes formatos de forma programática, para facilitar a integração com outras plataformas, como sistemas de legendagem de vídeos. No entanto, esta funcionalidade não foi implementada na fase atual, dado que o foco primário do STAAR é a criação do Diário da Assembleia da República, que requer um trabalho meticuloso de edição das transcrições. A implementação futura desta API poderá expandir a utilidade e a interoperabilidade do STAAR, permitindo uma maior integração com outras plataformas e sistemas.

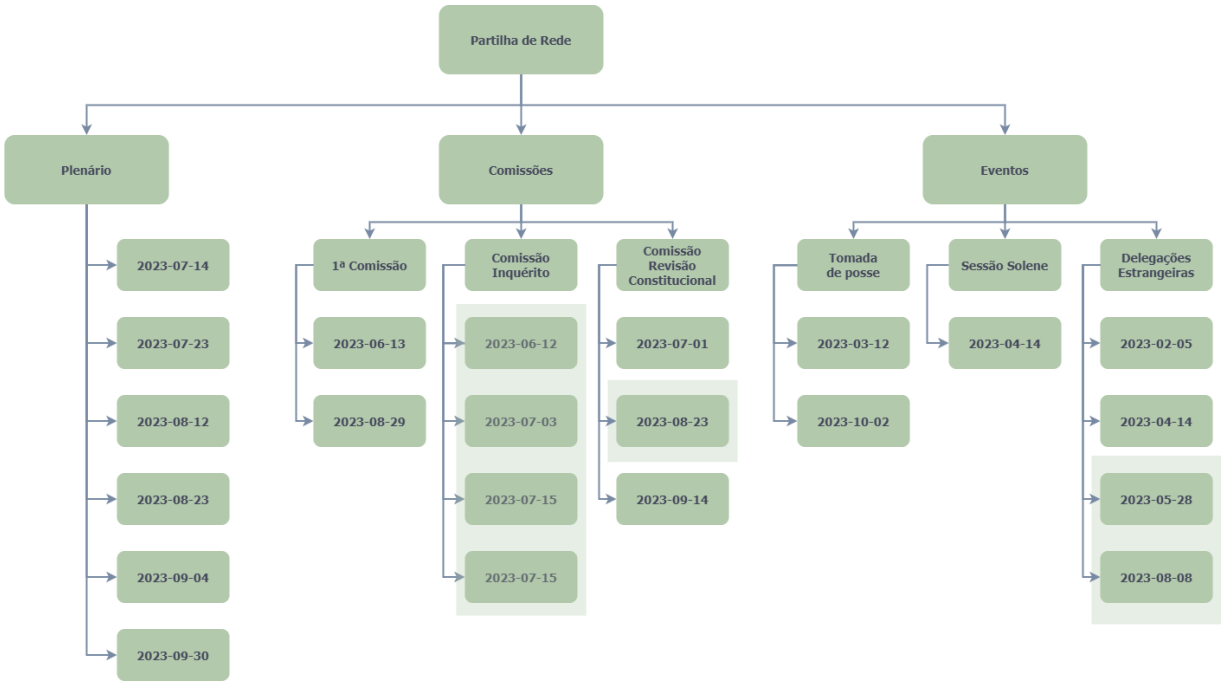


Figura 16 - Estrutura de pastas do repositório de transcrições

No que se refere à base de dados, o desenvolvimento foi efetuado de forma garantir uma representação precisa e completa das transcrições, bem como dos metadados associados. A escolha do SQL Server como motor da base de dados proporciona um ambiente robusto e confiável para o armazenamento e manipulação dos dados, e tirar ainda partido dos recursos existentes. A estrutura

da base de dados foi desenhada para facilitar consultas eficientes e fornecer uma visão clara do processo de transcrição, desde a recolha do áudio até a geração do texto transcrito.

A tabela 'transcricoes' foi criada para armazenar as informações relacionadas a cada transcrição realizada pelo STAAR. Os campos desta tabela e o seu tipo são descritos na Tabela 6.

Tabela 6 - Campos da base de dados de transcrições realizadas pelo STAAR

Campo	Tipo	Descrição
audio_ficheiro_nome	NVARCHAR	nome do ficheiro de áudio que foi transcrito
audio_ficheiro_caminho	NVARCHAR	caminho onde o ficheiro de áudio foi armazenado
audio_data_criacao	DATETIME	data e hora de criação do ficheiro de áudio
audio_duracao	INT	duração do ficheiro de áudio, em segundos
transcricao_ficheiro_nome	NVARCHAR	nome do ficheiro de transcrição gerado
transcricao_ficheiro_caminho	NVARCHAR	caminho onde o ficheiro de transcrição foi armazenado
processo_tempo_inicio	DATETIME	data e hora em que o processo foi iniciado
processo_tempo_fim	DATETIME	data e hora em que o processo foi concluído
processo_duracao	INT	duração do processo, em segundos
transcricao_duracao	INT	duração da etapa de transcrição, em segundos
alinhamento_duracao	INT	duração da etapa de alinhamento, em segundos
diarizacao_duracao	INT	duração da etapa de diarização, em segundos
contagem_palavras	INT	número de palavras presentes no ficheiro transcrito
nome_host	NVARCHAR	nome do equipamento que realizou a transcrição
versao	NVARCHAR	versão da aplicação que efetuou a transcrição
org	NVARCHAR	organismo a que a transcrição se refere

Ao longo do processo de transcrição, os metadados associados ao áudio e à transcrição vão sendo recolhidos. No final do processo de transcrição, um script em Python (Figura 37) é executado para inserir estas informações numa tabela em base de dados. Este script é responsável por formatar os dados conforme necessário e inseri-los como um novo registo na tabela 'transcricoes'.

3.4. Infraestrutura e recursos operacionais

O sucesso de qualquer sistema tecnológico não se baseia apenas na sua conceção e desenvolvimento, mas também na robustez e eficiência da infraestrutura que o suporta. Nesta secção são descritos os detalhes técnicos e operacionais que formam a espinha dorsal do Sistema de Transcrição Automática para a Assembleia da República (STAAR).

A Figura 17 ilustra de forma esquemática a interação e o fluxo de informação entre as componentes fundamentais do STAAR, o que permite obter uma visualização clara da estrutura e funcionamento das várias componentes.

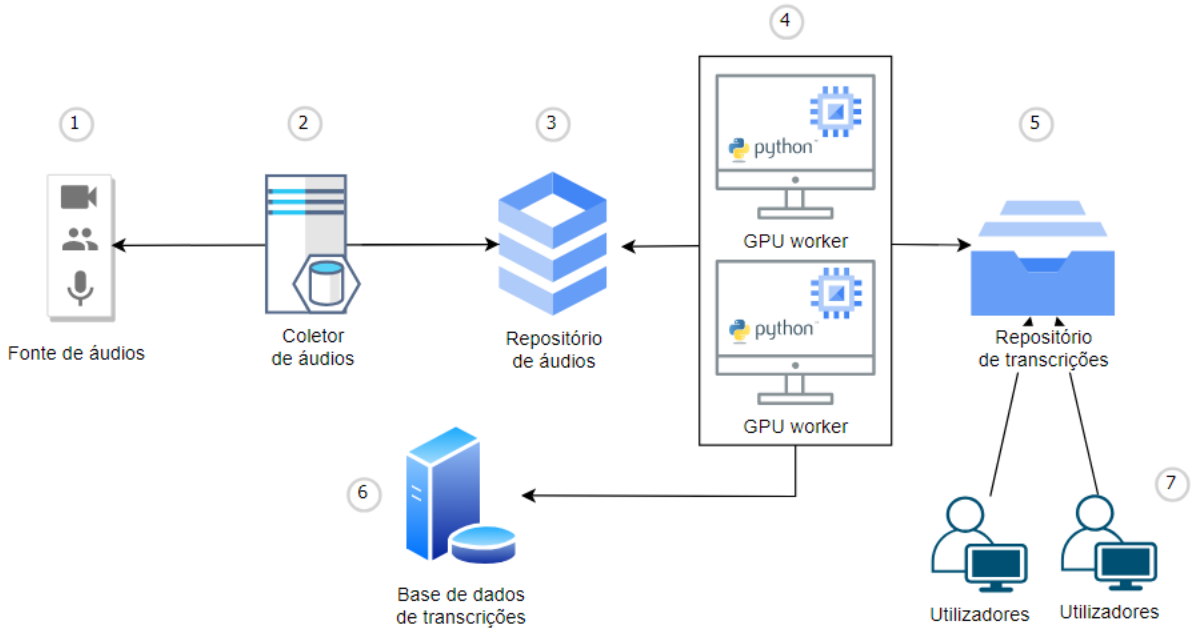


Figura 17 - Diagrama de recursos envolvidos no processo de transcrição

As componentes envolvidas no processo de transcrição são:

1. Fonte de áudios - Embora seja externa ao STAAR, é a partir daqui que todos os áudios, provenientes de diversas origens, são inicialmente obtidos. A seleção criteriosa destes áudios é fundamental para garantir que apenas os áudios pretendidos são transcritos;
2. Coletor de áudios - O Coletor de Áudios é um módulo especializado que interage diretamente com a Fonte de Áudios. A sua principal função é filtrar e selecionar os áudios pertinentes para transcrição, assegurando que apenas os áudios relevantes são processados, de acordo com o descrito em Recolha de áudios (3.3.1);
3. Repositório de Áudios - Este servidor de ficheiros é dedicado ao armazenamento dos áudios selecionados para transcrição. Equipado com capacidade de armazenamento robusta e escalável, garante que todos os áudios estão disponíveis para processamento;
4. GPU Worker - Uma infraestrutura baseada em GPU permite a utilização eficiente dos modelos de *deep learning*, que estão na raiz do sistema de Reconhecimento Automático de Fala (ASR) do Whisper. Este tipo de hardware avançado não só acelera o processo de transcrição, mas também garante uma análise precisa e eficaz dos áudios, mesmo em cenários de alta procura, como é o caso da transcrição de debates parlamentares, onde a rapidez e a precisão são essenciais. Portanto, a integração de GPUs é uma decisão estratégica que potencializa a eficiência e a eficácia do processamento de áudios pelo Whisper, que contribui significativamente para a qualidade e a rapidez das transcrições geradas. Neste âmbito foram usadas GPUs GeForce RTX 3060 com 12GB GDDR6, que a AR já dispunha. Estas placas GPU,

robustas e de alta performance, provaram ser bastante eficientes, mesmo ao lidar com os módulos mais exigentes do Whisper. Para otimizar o processo de transcrição e garantir um paralelismo eficaz, foi desenvolvido código em python de forma a coordenar a transcrição de áudios entre os diferentes postos, evitando duplicações e otimizar a utilização dos recursos disponíveis. Assim, estes equipamentos desempenham as tarefas de Processamento de áudio (3.3.2) e Tratamento de texto (3.3.3). A sua localização na infraestrutura da AR garante a conformidade com o requisito #15, ao garantir que todo o processamento é realizado internamente, sem recurso a *clouds* públicas;

5. Repositório de transcrições - Após a transcrição, os documentos resultantes são armazenados neste servidor de ficheiros. Com a partilha de ficheiros através da rede e mecanismos de controle de acesso rigorosos, garante-se que apenas os utilizadores autorizados acedem às transcrições, cumprindo assim o requisito #17;
6. Base de dados de transcrições - Este motor de base de dados, já existente na infraestrutura da AR, é responsável por armazenar todas as informações relevantes recolhidas durante o processo de transcrição. Funciona como um registo centralizado, que facilita a gestão e análise dos ficheiros transcritos. Caso se decida avançar no futuro com APIs de interligação do STAAR a outros sistemas, esta base de dados será vital para o efeito;
7. Utilizadores - Os utilizadores acedem às transcrições do repositório de transcrições através da rede e trabalham com as das ferramentas de produtividade da AR já existentes nos seus postos de trabalho, cumprindo assim o requisito #5.

CAPÍTULO 4.

Implementação e Resultados

Neste capítulo, é descrita a fase prática da implementação do Sistema de Transcrição Automática (STAAR) no ambiente parlamentar, com destaque para as adaptações e resultados obtidos, os desafios experienciados e como foram ultrapassados. A transição de um sistema teórico e concetual para uma solução funcional no mundo real exige uma série de considerações e ajustes. A implementação não se trata apenas de instalar e configurar software, mas também de garantir que a solução responde às necessidades específicas dos utilizadores e se integre harmoniosamente ao ambiente de trabalho existente. É ainda fundamental avaliar o desempenho do sistema em condições reais e recolher a opinião dos utilizadores para garantir que a solução é não só tecnicamente sólida, mas também prática e eficiente. Ao longo deste capítulo, são desenvolvidas cada uma dessas etapas, de forma a proporcionar uma visão abrangente do processo de implementação e dos resultados alcançados com o STAAR.

4.1. Fases de implementação

O STAAR foi desenvolvido de acordo com a metodologia DSRM (*Design Science Research Methodology*) [6]. Esta metodologia foi escolhida devido à sua abordagem iterativa e centrada no desenho, que se alinha perfeitamente com os objetivos de desenvolver uma solução adaptada às necessidades do ambiente parlamentar. Ao longo deste processo, ilustrado através da Figura 18, foram executadas três iterações distintas para refinar e otimizar a solução final.

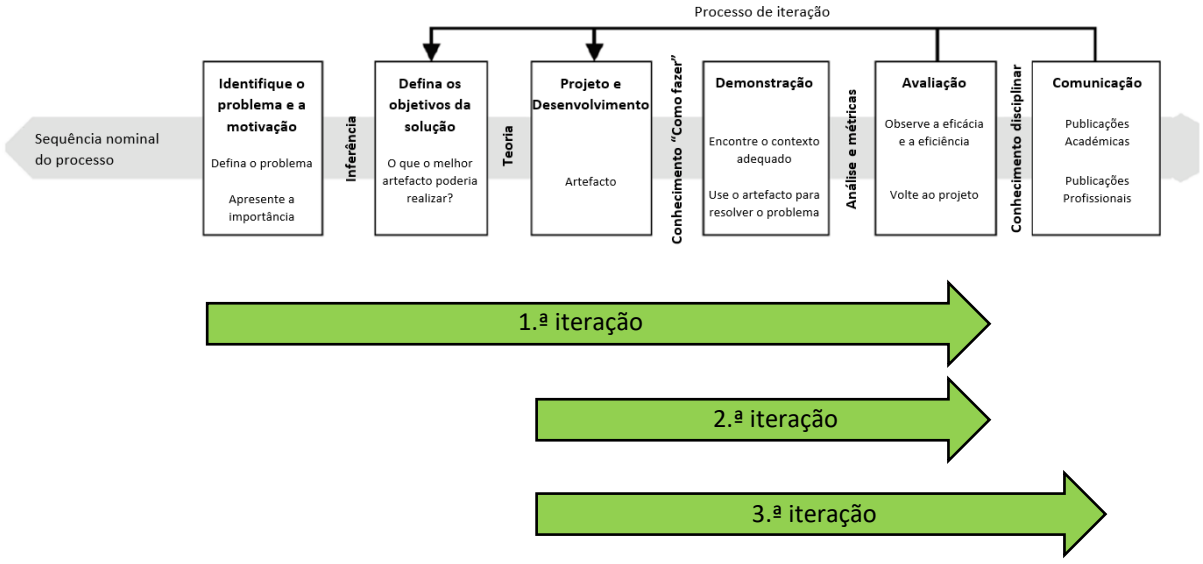


Figura 18 - Ciclos de iteração DSRM

Na primeira iteração, de dezembro de 2022 a março de 2023, foram identificadas as necessidades e estabelecidos os objetivos, que resultou no desenvolvimento de um protótipo do STAAR como forma de avaliar sumariamente se um modelo de inteligência artificial, como o Whisper, tinha a qualidade mínima que justificasse o investimento neste tipo de tecnologia.

Para a criação deste protótipo, optou-se por utilizar o Google Colab [62], uma plataforma em *cloud* amplamente reconhecida no mundo da ciência de dados e aprendizagem de máquina. A sua natureza intuitiva, juntamente com o facto de ser gratuito e não necessitar de configuração, tornou-o uma escolha ideal para uma prova de conceito. Além disso, a capacidade de aceder a recursos computacionais avançados, como GPUs, sem qualquer custo, foi uma vantagem significativa.

A escolha de realizar esta demonstração numa plataforma em *cloud*, sem a necessidade de configurações complexas ou investimentos iniciais, alinha-se com a filosofia do DSRM de criar soluções práticas e orientadas para o utilizador. Esta demonstração não só permitiu validar a ideia central do STAAR, mas também fornecer informações valiosas para as iterações subsequentes do design e desenvolvimento do sistema.

Através da documentação disponível do Whisper, foi desenvolvido um *script* que os utilizadores executavam na *cloud* para fazer transcrições quando necessitavam (Figura 19). Após a inicialização do sistema no Google Colab, os utilizadores tinham de fazer *upload* do ficheiro de áudio que pretendiam transcrever e, com o modelo de linguagem da Whisper, era produzido um ficheiro de texto com a transcrição integral do áudio submetido. O ficheiro era depois era descarregado e aberto no posto de trabalho do utilizador para servir como base à tarefa de transcrição de um debate parlamentar.

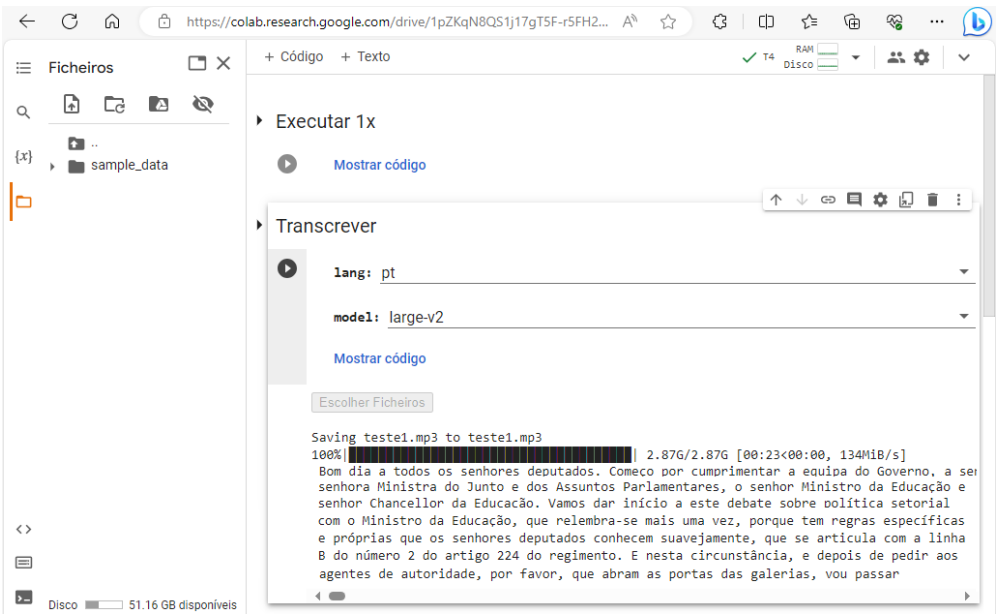


Figura 19 - Exemplo de transcrição de um áudio no Google Colab

Na segunda iteração, com base nas informações e *feedbacks* obtidos, o foco passou a ser a otimização e automação do processo de transcrição, tendo decorrido entre abril e agosto de 2023.

Em primeiro lugar foi implementado um sistema automatizado para a recolha dos áudios, de acordo com a arquitetura planeada. Em vez de depender de uploads manuais, o novo sistema foi projetado para identificar e processar áudios de forma mais fluida, reduzindo a intervenção manual e, consequentemente, o potencial de erros humanos.

A transcrição, uma vez mais realizada pelo modelo Whisper, foi otimizada para garantir maior precisão e rapidez usando recursos existentes na infraestrutura da AR, nomeadamente com GPUs dedicadas para o efeito. O sistema foi configurado para reconhecer e transcrever áudios de forma mais eficiente para aproveitar ao máximo as capacidades do modelo de linguagem.

Outro avanço significativo desta fase foi a introdução de um léxico personalizado. Este léxico, criado especificamente para o contexto do projeto, contem termos, expressões e nomenclaturas frequentemente usadas em debates parlamentares e que não eram corretamente transcritas. Ao integrar este léxico no processo de transcrição, o sistema passou a substituir automaticamente palavras transcritas de forma imprecisa por versões corretas, o que não só melhorou a precisão das transcrições, mas também economizou tempo, uma vez que eliminou a necessidade de correções manuais extensas.

Por último, para facilitar a acessibilidade e a utilização das transcrições, o sistema foi configurado para disponibilizar os textos transcritos no formato docx. Este formato permitiu aos utilizadores aceder, editar e partilhar as transcrições com facilidade, através de uma integração simples com ferramentas de processamento de texto já em uso na Assembleia da República.

Assim, a segunda iteração do DSRM representou um salto qualitativo no desenvolvimento do STAAR ao tornar o processo de transcrição mais ágil, preciso e alinhado às necessidades específicas do contexto parlamentar.

A terceira iteração, implementada em setembro de 2023, consistiu numa fase de refinamento e inovação, que permitiu alcançar avanços tecnológicos significativos que elevaram a qualidade e eficiência do STAAR a um novo patamar.

Um dos marcos desta fase foi a introdução do WhisperX, uma versão avançada e otimizada do modelo Whisper. Com a nova versão o tempo de transcrição foi drasticamente reduzido, o que tornou o processo não apenas mais rápido, mas também mais eficiente em termos de recursos computacionais. Esta melhoria foi crucial para garantir que o sistema conseguisse lidar com volumes de áudio maiores sem comprometer a rapidez ou a qualidade da transcrição.

Outra inovação significativa foi a implementação do alinhamento dos *timestamps* palavra a palavra. Isto permitiu que cada palavra transcrita fosse associada ao seu momento exato no áudio original, o que facilita a revisão e a correção de eventuais imprecisões.

A diarização do texto por orador foi outro avanço notável. O sistema foi melhorado para identificar e distinguir diferentes oradores num áudio, de forma a organizar a transcrição em que cada intervenção é claramente atribuída ao seu respetivo orador. Esta funcionalidade não só melhorou a legibilidade das transcrições, mas também facilitou a compreensão do contexto e da dinâmica das discussões.

Por fim, o léxico utilizado na fase anterior foi revisto e expandido para responder às lacunas entretanto identificadas.

4.2. Demonstração do sistema

A fase de demonstração é uma das etapas da metodologia DSRM [6], onde o artefacto (neste caso, o sistema de transcrição automática STAAR) é testado em condições reais para validar a sua eficácia e utilidade. Esta fase é planeada para envolver os principais *stakeholders* e garantir que a solução desenvolvida responde às necessidades elencadas na fase inicial do DSRM.

No final de 2022 foi formalmente nomeado um grupo de trabalho, Grupo de Trabalho da Transcrição Automática (GT-TA), com o objetivo de avaliar uma solução de transcrição comercial e composto por uma variedade de profissionais com competências e responsabilidades distintas, mas todos com um interesse comum na transcrição automática:

- 7 elementos da Divisão de Redação (DR) - elementos da equipa que realiza a transcrição dos debates e a produção do Diário da Assembleia da República (DAR);
- 1 elemento da Divisão de Apoio às Comissões (DAC) - embora a sua função principal não seja a transcrição de reuniões de comissão, é da sua responsabilidade de criar sumários das reuniões e pontualmente transcrever determinadas intervenções relevantes;
- 1 elemento do Canal Parlamento - sendo o serviço responsável pela gravação e disponibilização dos áudios das reuniões de plenário e de comissões, o papel deste elemento no grupo de trabalho foi fundamental garantindo que os aspetos técnicos relacionados com esta matéria fossem devidamente considerados;
- 1 elemento da Divisão de Infraestruturas Tecnológicas (DIT) – o autor do presente trabalho de projeto e responsável por recolher as informações necessárias e promover o desenvolvimento de uma solução de transcrição;
- Chefias dos diversos serviços interessados na transcrição automática.

Após resultados insatisfatórios com o produto comercial sugerido para análise, especialmente devido à falta de uma interface adequada para a manipulação do texto transcrito, ao tempo excessivo para a transcrição dos áudios e, principalmente, à elevada taxa de erro por palavra (WER) que se situou nos 39%, o GT-TA decidiu explorar outras alternativas.

Além do produto inicialmente proposto, entre janeiro de 2022 e fevereiro de 2023, a equipa de transcrição da AR testou cinco outros produtos, avaliando-os com base na WER. Conforme indicado num relatório interno elaborado por esta equipa, houve uma variação significativa nos resultados com valores de WER a oscilar entre os 23,1% e 44,4%.

Assim, foi no âmbito deste grupo de trabalho que se decidiu realizar as primeiras demonstrações baseadas no modelo de inteligência artificial Whisper e prosseguir com o desenvolvimento da solução caso os testes iniciais se revelassem interessantes em termos de WER e facilidade de uso.

Após o desenvolvimento refente à primeira iteração DSRM a equipa de trabalho GT-TA, durante o mês de fevereiro 2023, usou o Google Colab como plataforma de transcrição, aproveitando a sua acessibilidade e capacidade de processamento. Este período de utilização serviu como uma fase de avaliação, permitindo à equipa testar transcrições realizadas pelo modelo Whisper em diferentes cenários e com diferentes tipos de áudios. Ao transcrever áudios com características tão variadas, a equipa teve a oportunidade de avaliar a precisão e eficácia do Whisper em diferentes contextos, desde debates mais formais até intervenções mais espontâneas e dinâmicas. Visto a transcrição de áudios ter sido realizada de forma individual pelos vários elementos do GT-TA não é possível determinar com exatidão a quantidade de áudios processados.

No mês seguinte, em abril de 2023, a versão do STAAR relativa à segunda iteração do DSRM foi disponibilizada. Esta atualização permitiu que as transcrições fossem executadas automaticamente assim que os áudios estavam disponíveis internamente, utilizando os recursos internos da AR. O resultado era um documento no formato docx. Durante os meses de abril a julho de 2023, toda a equipa de transcrição da Divisão de Redação da AR passou a utilizar e avaliar as transcrições geradas de forma automática pelo STAAR, tendo sido processados 1915 ficheiros, o que equivale a aproximadamente 500 horas de áudio.

Após a análise da avaliação obtida da segunda iteração, foram implementados novos desenvolvimentos com o objetivo de colmatar as lacunas identificadas. Os detalhes dos problemas identificados e das novas funcionalidades para os endereçar estão descritos na secção 4.3.3.

Em setembro de 2023, uma versão atualizada do STAAR foi disponibilizada, correspondendo à terceira e última iteração do ciclo DSRM. Esta nova versão foi submetida a testes e avaliações pela equipa completa de transcrição da Divisão de Redação da AR. Nesta última iteração, entre setembro e final de outubro de 2023, foram transcritos 486 ficheiros, o que representa cerca de 225 horas de arquivo áudio.

4.3. Avaliação

Através da avaliação do Sistema de Transcrição Automática (STAAR) é possível medir eficácia e precisão do sistema no contexto dos debates parlamentares. Esta avaliação compreende métricas técnicas de transcrição, taxas de erro (WER) bem como a avaliação mais holística através da metodologia adotada (DSRM).

4.3.1. Métricas de transcrição

Para poder avaliar a eficácia e precisão do Sistema de Transcrição Automática da Assembleia da República (STAAR), é essencial quantificar métricas resultantes das transcrições realizadas. Nesta secção, são apresentadas estatísticas detalhadas referentes ao volume de reuniões plenárias e comissões processadas pelo STAAR, bem como as horas de áudio associadas e palavras transcritas.

É de ressalvar que a primeira iteração não tem dados estatísticos uma vez cada utilizador recorria à plataforma Google Colab sempre que pretendia testar uma transcrição. No entanto os dados desta iteração não são muito relevantes na medida em que, para além de serem poucos, não foram de facto realizados pelo STAAR, mas sim por uma plataforma externa, ainda que usasse o mesmo modelo de IA para a transcrição, o Whisper.

Conforme se pode avaliar pela Tabela 7, desde início de abril até o final de outubro de 2023, foram transcritas de forma automática pelo sistema mais de 30 dias de áudio contínuo (724 horas), referentes a 335 reuniões plenárias, comissões e eventos.

Tabela 7 - Métricas de transcrição

Iteração	Modelo	Período	Órgão	Reuniões	Horas	Palavras
2ª	Whisper	abril a agosto	Plenário	50	180	1.495.826
			Comissões	122	318	2.537.262
			Eventos	2	1	11.000
3ª	WhisperX	setembro a novembro	Plenário	22	79	675.894
			Comissões	139	146	1.053.725
Total				335	724	5.773.707

A Figura 20 permite avaliar o total de horas de áudio transcritas pelo STAAR, por tipo de reunião, durante os meses de abril a de outubro de 2023.

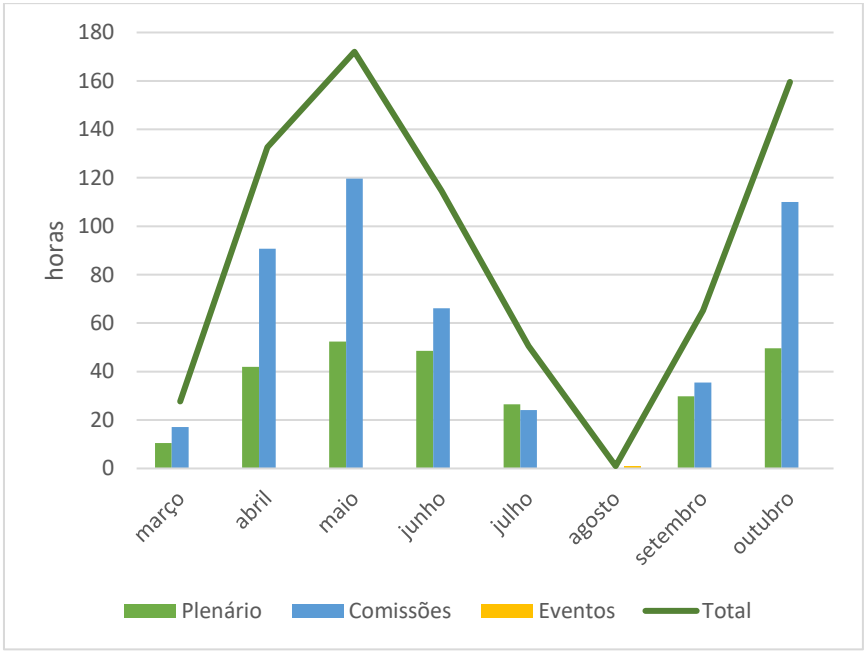


Figura 20 - Total de horas de áudio transcritas

De acordo com o descrito na seção 4.1, Fases de implementação, nas duas primeiras iterações o STAAR apenas transcrevia áudio para texto, sendo que na terceira iteração passou também a alinhar os *timestamps* do texto e a realizar diarização. O processo de transcrição passou a ser efetuado pelo WhisperX que recorre ao modelo Whisper disponibilizado pela OpenAI, mas com uma grande redução a nível de consumo de recursos de GPU e tempo necessário para a transcrição.

Conforme se pode ver pela Figura 21, a etapa de transcrição de um fragmento de 15 minutos de uma reunião plenária, passou em média de 3 minutos e 45 segundos (225 segundos) para apenas 33 segundos, ou seja, um aumento de eficiência em cerca de 86%.

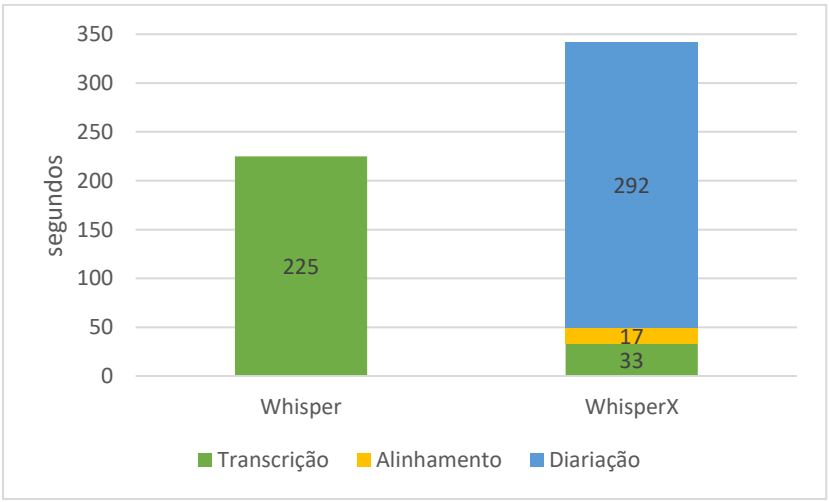


Figura 21 - Comparação de tempo total de processamento entre Whisper e WhisperX

No entanto, e como se pode observar, a duração global do processo de transcrição aumentou da segunda iteração (225 segundos) para a terceira iteração (342 segundos). Este aumento deve-se ao facto de, para além da transcrição, na terceira iteração serem realizadas as tarefas de alinhamento e diarização. Embora o processo de alinhamento seja muito rápido (17 segundos), o processo de diarização é bastante mais demorado (292 segundos), o que representa cerca de 85% do tempo total da etapa de Processamento de áudio. Apesar do tempo total de transcrição ter aumentado, mas mantendo-se bastante inferior aos valores considerados como razoáveis (requisito #4), o valor acrescentado no texto final, proporcionado pela diarização, justifica este acréscimo especialmente no contexto dos debates parlamentares.

4.3.2. Taxa de erro das transcrições

Para uma análise objetiva e quantitativa da taxa de erro do sistema de transcrição automática, é essencial começar com uma transcrição manual. As transcrições manuais foram realizadas de forma a refletir fielmente o discurso proferido, sem edições ou adições de informações contextuais, como aplausos ou protestos.

Esta abordagem procurou evitar possíveis enviesamentos que pudessem ser introduzidos por transcrições automáticas pré-existentes. Em avaliações anteriores, por questões de tempo e para obter uma WER que refletisse as edições feitas pela equipa de transcrição da AR, foram usadas transcrições já editadas ou até publicadas. Embora esta abordagem possa ter introduzido algumas distorções na WER, a metodologia consistente e a quantidade significativa de transcrições analisadas garantem que as comparações sejam válidas.

Observou-se que as transcrições manuais do mesmo discurso, realizadas por diferentes pessoas, apresentavam variações. Estas diferenças podem ser atribuídas a um "enviesamento profissional", onde quem realiza a transcrição pode optar por ignorar certas palavras, como hesitações ou repetições. Além disso, os critérios de edição individuais, como a utilização de pontuação específica ou a transcrição de palavras incompletas, também influenciaram as transcrições.

Para minimizar estas discrepâncias, decidiu-se padronizar certos aspetos das transcrições manuais como por exemplo, palavras mal pronunciadas, hesitações, repetições e correções feitas pelo orador foram mantidas nas transcrições.

No Anexo II - Análise comparativa de transcrições é apresentado um discurso específico cujo WER é de 2%, conforme se pode constatar pela Figura 22. O anexo inclui a transcrição manual, a gerada pelo STAAR, bem como uma comparação entre elas, onde são destacadas as diferenças identificadas.

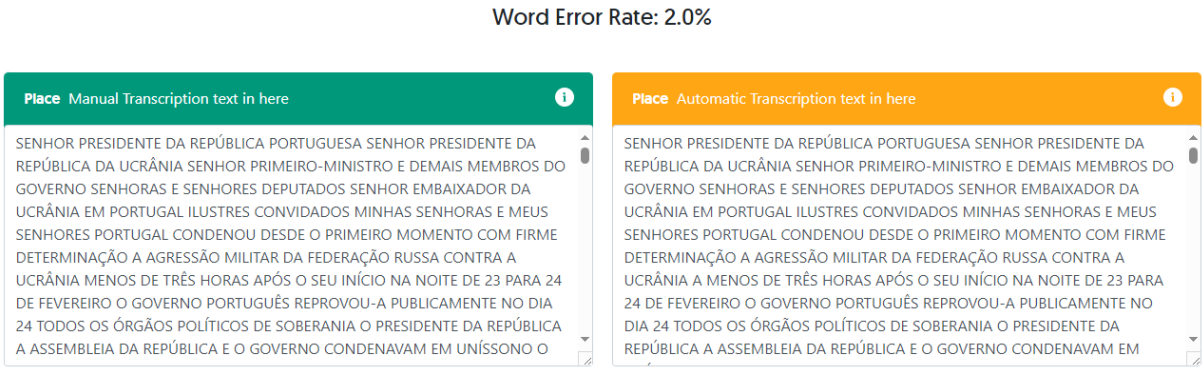


Figura 22 - Análise de WER de transcrições, através do site Amberscript [9]

4.3.2.1. Transcrições avaliadas

Nesta secção é feita uma análise detalhada das transcrições para avaliar a precisão do STAAR. O objetivo é examinar um conjunto de transcrições que reflitam as variadas circunstâncias e fatores que habitualmente influenciam a qualidade da transcrição. Esta análise inclui tanto fragmentos de áudio completos com múltiplos oradores como discursos individuais.

Foi dada especial atenção a diferentes estilos de discurso, seja ele pré-redigido e lido ou improvisado, e a diferentes contextos parlamentares frequentemente transcritos, como sessões plenárias e reuniões de Comissões Parlamentares. Também se considerou o ambiente do debate, que varia de discussões calmas, sem interrupções, a debates mais intensos com ruído de fundo e várias interrupções.

Assim, foram analisados três discursos/intervenções lidas em Plenário e um depoimento lido na Comissão Parlamentar de Inquérito à Tutela Política da Gestão da TAP (CPI TAP), cinco intervenções de improviso em Plenário e duas curtas intervenções de improviso em Comissão Eventual para a Revisão Constitucional (CERC), para tentar avaliar a influência de cada orador/discurso na qualidade das transcrições automáticas efetuadas.

O objetivo é entender como características individuais do orador, como pronúncia, ritmo, entoação e clareza de fala, influenciam a qualidade da transcrição. Esta análise incluiu oradores de diferentes regiões de Portugal e até um caso com possível distúrbio de fala, conforme se pode observar na coluna “Pronúncia” da Tabela 8.

Além disso, para avaliar o efeito de interrupções, mudanças de orador, ruído de fundo e comentários paralelos na qualidade das transcrições, foram analisados fragmentos de várias reuniões e debates. Estes incluíram debates acalorados com várias interrupções e ruído de fundo.

No total, 11 transcrições foram analisadas, totalizando pouco mais de duas horas. Apesar da duração relativamente curta, considera-se que estas transcrições oferecem uma visão representativa

dos desafios enfrentados diariamente pela equipa que realiza a transcrição de debates parlamentares. Uma síntese desta análise pode ser encontrada na Tabela 8.

Tabela 8 - Análise de transcrições e WER do STAAR

#	Órgão	Tipo de intervenção	Pronúncia	Tipo de debate	WER por tipo de transcrição (%)	
					Manual / Fonética	Sem disfluências
T1	Plenário	Lida / Int. Escrita	Norte	Sem ruído ou interrupções	2,0	
T2	Plenário	Improviso	Centro	Ruidoso e com interrupções	9,7	4,5
T3	Plenário	Lida / Int. Escrita	Centro	Sem ruído ou interrupções	1,7	
T4	Plenário	Improviso	Centro	Ruidoso e com muitas interrupções	8,2	
T5	Plenário	Lida/Int. Escrita	Madeira	Sem ruído ou interrupções	8,1	
T6	Plenário	Improviso	Centro / Norte	Ruidoso e com muitas interrupções	5,8	1,8
T7	CPI	Apresentação / Depoimento lido	Dislália	Ruidoso	6,1	5,4
T8	CPI	Depoimento lido / Pergunta / Resposta	Dislália	Ruidoso e com interrupções	7,3	
T9	CPI	Pergunta /Resposta	Dislália	Ruidoso e com interrupções	11,3	
T10	CERC	Improviso	Centro	Ruidoso	5,2	
T11	CERC	Improviso	Centro	Ruidoso	4,0	1,5

4.3.2.2. Erros comuns do Whisper

Para identificar os erros mais comuns na transcrição do Whisper em condições ideais, que não são influenciados por fatores externos, realizou-se uma análise detalhada da transcrição presente no Anexo II - Análise comparativa de transcrições, um discurso proferido pelo Presidente da Assembleia da República durante uma sessão solene. Esta análise encontra-se no Anexo III - Exemplos de erros comuns em transcrições com Whisper, onde foram utilizadas as transcrições já corrigidas pela equipa de transcrição, e não as versões uniformizadas usadas para calcular a taxa WER.

O primeiro aspeto a destacar desta análise é a quantidade relativamente baixa de correções necessárias na transcrição do Whisper (aproximadamente 40, que variam conforme se contabilizam palavras individuais ou expressões) quando comparada à transcrição manual.

Destas, apenas uma pequena fração dos erros pode ser claramente atribuída a falhas básicas de transcrição, onde palavras foram transcritas de forma errada, mesmo sendo facilmente reconhecíveis ao ouvido humano, como se pode verificar pelas linhas 1, 3, 13, 18, 21, 26 e 29 do referido anexo. Um erro específico, mencionado na linha 18, "solenamente" em vez de "solenemente", pode ter ocorrido devido à frequente grafia incorreta dessa palavra em muitos conteúdos online, mas é inegavelmente um erro ortográfico.

Embora alguns erros possam ser atribuídos à pronúncia ou dicção do orador, os erros nas linhas 2 e 4 também se enquadram nessa categoria. Acresce que, dois erros parecem estar alinhados com a ortografia do português brasileiro (como "excepcional" na linha 7 e "conduzente" na linha 31). Embora necessitem de correção, é complexo categorizá-los simplesmente como erros de transcrição, sendo mais apropriado considerá-los como desvios do AO90 [63].

Os restantes erros identificados podem ser categorizados em quatro principais grupos:

- Nomes próprios (12 ocorrências) - referentes às linhas 16, 20, 28, 30, 31, 32, 33 e 34. Para esta contagem, considerou-se um único erro quando duas ou mais palavras consecutivas estavam erradas. Muitos dos erros são relacionados a nomes estrangeiros, especialmente ucranianos, cuja adaptação para o alfabeto latino não é uniforme, o que pode complicar a tarefa do Whisper em transcrevê-los de forma precisa;
- Princípios e convenções (8 ocorrências) - referentes às linhas 5, 15, 17, 19, 23, 25, 27 e 36. A maioria destes erros está relacionada com a aplicação inadequada de maiúsculas, minúsculas ou abreviaturas, conforme estabelecido no guia de elaboração do DAR, um documento usado pela equipa de transcrição com regras de escrita;
- Pronúncia (7 ocorrências) - referentes às linhas 8, 9, 10 e 12. Embora a transcrição do Whisper esteja correta, os verbos no pretérito perfeito do indicativo deveriam ter um acento na sílaba tónica "a" para se diferenciarem do presente do Indicativo, necessitando, assim, de correção;
- Disfluência (3 ocorrências) - referentes às linhas 14, 20 e 24. Nestas situações, o Whisper corrigiu repetições, hesitações ou palavras mal pronunciadas, eliminando a necessidade de correção manual, caso a transcrição fosse realizada *ipsis verbis*. Isto indica que, em certos contextos, a WER calculada poderia ser inferior, uma vez que vai de encontro ao trabalho que um humano também faria ao remover as disfluências.

4.3.2.3. Análise da WER

No contexto das transcrições analisadas, embora o número e a duração destas sejam limitados, o que sugere a necessidade de análises mais extensas e uma interpretação cautelosa dos resultados, é evidente que a WER do STAAR é, em média, significativamente inferior ao intervalo inicialmente antecipado e considerado pela AR como mínimo aceitável por um sistema de transcrição automático (entre 10% e 15%). Em situações ideais, como intervenções lidas em sessões plenárias, a WER chega a ser inferior a 4,3%, o valor que a própria OpenAI, criadora do modelo de linguagem Whisper utilizado pelo STAAR, indicou para transcrições em português, conforme ilustrado na Figura 23.

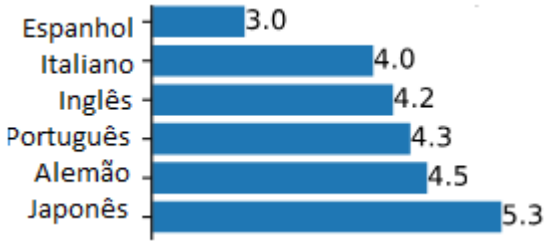


Figura 23 - WER do modelo large-v2 do Whisper para a transcrição em várias línguas [50]

De facto, uma análise preliminar dos resultados indica que a transcrição de discursos preparados e lidos é, em geral, mais precisa do que a de discursos espontâneos. Similarmente, as intervenções no Plenário tendem a ser transcritas com maior precisão do que as realizadas em reuniões de Comissão. De acordo com os dados obtidos, sintetizados na Tabela 8, o menor valor de WER observado foi de 1,7% numa intervenção lida da tribuna (T3), seguido de 2% num discurso do Presidente da AR numa sessão solene (T1). Por outro lado, o valor mais elevado de WER (11,3%) foi observado na transcrição de uma sessão de perguntas e respostas na Comissão de Inquérito da TAP (T9).

Por outro lado, o testemunho lido por um dos mesmos oradores na T7 regista a WER mais baixa (6,1%) de todas as transcrições feitas em Comissão. Este facto parece reforçar a noção de que, mantendo todos os outros fatores constantes, a natureza da intervenção, se lida ou improvisada é o principal fator externo ao STAAR que impacta a qualidade da transcrição.

No caso das intervenções pré-escritas e lidas, a maior precisão pode ser atribuída tanto a fatores fonéticos, por exemplo uma dicção mais clara e uma prosódia aprimorada, quanto à redução de hesitações, repetições ou correções feitas pelos oradores em tempo real.

Além disso, fatores ambientais e regimentais também desempenham um papel. Intervenções escritas, como as mencionadas em Plenário (T1 e T3), são frequentemente proferidas da tribuna ou da Mesa da Assembleia da República, na Sala das Sessões. Estas, por natureza, enfrentam menos interrupções e o seu áudio é menos perturbado pelo ruído ambiente. Esta observação também se

aplica ao testemunho lido na CPI TAP na T7, apesar das peculiaridades na fala do orador e do ambiente geralmente mais barulhento das reuniões de Comissão, devido à proximidade entre oradores, público, assessores políticos, funcionários parlamentares e jornalistas presentes nas salas das comissões.

Naturalmente, as transcrições de intervenções de improviso revelam uma WER consideravelmente mais alta. Isso ocorre não só devido às características típicas do discurso espontâneo (como hesitações, palavras truncadas, repetições ou autocorreções), mas também devido à sua natureza regimental intrínseca. Estas intervenções frequentemente enfrentam protestos e interrupções que levam muitas vezes os oradores a repetir, ajustar ou enfatizar partes do seu discurso, seja para retomar o ponto após a interrupção ou para responder diretamente às interjeições.

A natureza da intervenção ou debate, especialmente quando consideramos o nível de ruído de fundo e as interrupções, combinado com a decisão de ler ou improvisar a intervenção, parece ser o segundo fator mais influente na qualidade da transcrição automática. Notavelmente, a transcrição de uma intervenção em Plenário com a WER mais elevada (9,7%) corresponde a um pedido de defesa da honra seguido de uma resposta (T2). Em contraste, numa reunião de Comissão, a transcrição com a WER mais alta (11,3%) foi, como mencionado anteriormente, de uma sessão de perguntas e respostas.

As particularidades individuais da fala de cada orador, como pronúncia, dicção e prosódia, parecem ter um impacto menor na qualidade da transcrição automática em comparação com os fatores anteriormente mencionados. De facto, as transcrições T1 (com um orador com pronúncia do norte de Portugal), T5 (uma deputada da região autónoma da Madeira), T7 (um orador com aparente dislália) e T10 (um orador com ritmo de fala acelerado) não exibem os valores mais elevados de WER observados.

No que diz respeito à influência de interrupções, mudanças de orador, ruído de fundo e outros apartes na qualidade das transcrições, observou-se um número limitado de erros, como a omissão de frases completas no início ou fim dos fragmentos e durante a mudança de oradores, que revelou a necessidade de afinar parâmetros técnicos relativos à deteção de fala (VAD).

Efetivamente, enquanto a transcrição T9, com a WER mais alta (11,3%), corresponde a um debate com frequentes mudanças de orador numa reunião de Comissão, a transcrição T6, de um fragmento do último debate do estado da Nação na sala das sessões, não apresentou a WER mais alta, apesar de ter um ambiente aceso e interrupções constantes. Este facto sugere que a proximidade de outros Deputados ao microfone do orador e o consequente ruído de fundo podem ter um impacto maior na qualidade da transcrição do que a simples alternância entre oradores.

Por fim, um dos elementos que se procurou minimizar ou eliminar nas transcrições manuais para obter um valor de WER mais preciso, como repetições, hesitações e frases iniciadas, mas não concluídas pelos oradores, parece influenciar consideravelmente o aumento desse valor. Contudo,

essa variação parece estar mais relacionada a uma omissão de repetições pelo Whisper do que aos próprios elementos em si.

De facto, ao excluir essas repetições da transcrição manual, como observado nas transcrições T2, T6, T7 e T11 (conforme indicado na coluna final da Tabela 8), e tornando a transcrição manual menos alinhada ao que foi realmente expresso pelo orador, nota-se uma redução significativa na taxa WER. Especificamente, nas T2 e T6 (transcrições em Plenário), a taxa WER diminui de 9,7% para 4,5% e de 5,8% para 1,8%, respetivamente.

Isto indica que a omissão intencional de repetições pelo Whisper, similar ao que um redator experiente faria na transcrição manual, mesmo elevando a WER, na verdade facilita o trabalho de transcrição, uma vez que evita a necessidade de transcrever grande quantidade de texto que, de acordo com as práticas e convenções da AR, teria de ser posteriormente removido manualmente do texto final.

Com base nos dados observados, é possível determinar quais os fatores que influenciam a qualidade da transcrição automática de debates parlamentares, e classificá-los com base no grau de impacto no WER, consolidados na Tabela 9.

Tabela 9 - Fatores que influenciam a transcrição e o grau de influência na WER

Fator	Breve descrição	Grau de influência na WER
Natureza da intervenção	Se a intervenção é pré-escrita e lida ou se é espontânea	Alto
Ambiente do debate	Nível de ruído e número de interrupções durante a fala	Alto
Mudança de orador/ Atribuição de Palavra	Frequência de mudança entre oradores ou interjeições	Médio
Omissão de disfluências	Se o sistema omite disfluências (repetição de palavras, palavras cortadas, etc.) deliberadamente	Médio
Características de fala do orador	Pronúncia, dicção, ritmo, etc.	Baixo
Características regimentais da intervenção	Natureza da intervenção, como pedidos de esclarecimento, etc.	Baixo

Apesar da dimensão da amostra testada não ser muito extensa e do curto intervalo de tempo de que se dispôs para fazer esta análise, a diversidade de discursos apreciados permite ainda assim

concluir, com alguma segurança, que a eficiência e a taxa de precisão do STAAR superam não só as de todos os softwares e soluções previamente testadas, como as próprias previsões iniciais do GT-TA em relação às capacidades de um sistema de transcrição automático.

Com efeito, nas melhores condições possíveis, a WER calculada ficou entre os 1,7% e os 2%, enquanto nas piores condições testadas, a WER não foi além dos 11,3 %, ficando, portanto, abaixo dos 15% inicialmente propostos como limite máximo aceitável pelo GT-TA (requisito #14 da Tabela 4).

4.3.3. Avaliação DSRM

Esta secção descreve a etapa de Avaliação do *Design Science Research Methodology* (DSRM) [6] que permite obter uma compreensão clara do desempenho do sistema no contexto real, identificar áreas de sucesso e potenciais melhorias e desta forma alinhar o desenvolvimento com os objetivos iniciais do projeto.

Após o período de demonstração do protótipo de transcrição automática com Whisper no Google Cloud, a primeira iteração, os utilizadores reconheceram utilidade no sistema, mas identificaram várias áreas que necessitavam de melhoria:

- Complexidade do processo - a transcrição, embora realizada automaticamente, era percecionada como um processo complexo e demorado, o que poderia comprometer a eficiência desejada;
- Tempo necessário à transcrição - por vezes o sistema em *cloud* demorava 10 minutos para transcrever um áudio de 15 minutos e falhava várias vezes, o que obrigava a repetir o processo;
- Dificuldade no manuseamento de ficheiros - a necessidade de fazer upload de cada áudio individualmente e, posteriormente, descarregar o documento de texto correspondente, tornava o processo trabalhoso e pouco prático, especialmente quando se tratava de um grande volume de áudios;
- Incorreções na transcrição - o sistema, na sua forma inicial, não estava adequadamente adaptado à terminologia e gíria específica do contexto parlamentar, o que resultava em erros de transcrição que, embora corrigíveis, exigiam um esforço adicional por parte dos revisores;
- Regras de escrita - existiam diversas regras de escrita e formatação que não eram aplicadas pela transcrição, o que exigia uma revisão manual para garantir a conformidade com os padrões estabelecidos.

A avaliação do resultado da segunda iteração foi feita pela mesma equipa, o GT-TA, tendo verificado que muitos dos problemas identificados tinham sido resolvidos. No entanto, e para dar

resposta integral aos requisitos identificados (Tabela 4), foi necessário melhorar o sistema com o objetivo de resolver as seguintes situações:

- Imprecisões temporais - os registos temporais associados às frases transcritas não eram precisos, o que poderia comprometer a fidelidade da transcrição em relação ao áudio original;
- Formato do texto - a ausência de quebras de linha no texto transcrito dificultava a leitura e a navegação pelo documento, tornando a revisão e edição uma tarefa mais desafiadora;
- Ausência de transcrição no início/fim de intervenções - constatou-se que por vezes o sistema não transcrevia segmentos que fossem próximos, tanto antes como após, a momentos de silêncio no áudio, muitas vezes correspondentes à mudança de orador.

A terceira iteração representou a fase final de desenvolvimento e otimização do sistema. Com base no *feedback* contínuo do GT-TA e nas lições aprendidas nas iterações anteriores, o desenvolvimento focou-se em melhorar as funcionalidades existentes e em introduzir novas características para melhor atender às necessidades dos utilizadores. Nesta fase, foram introduzidas melhorias significativas na precisão da transcrição, nos *timestamps* associados a cada palavra, na identificação do texto referente a cada orador, na formatação do texto e afinados parâmetros relativos à deteção de fala (VAD). Além disso, foram implementadas funcionalidades adicionais, como a capacidade de personalizar dicionários de substituição e a integração com outras ferramentas e plataformas utilizadas no ambiente parlamentar.

O envolvimento contínuo do GT-TA ao longo de todas as etapas do DSRM garantiu que o sistema desenvolvido estivesse alinhado com as expectativas e requisitos dos utilizadores finais. A Tabela 10 apresenta a avaliação realizada pelo GT-TA, com base nos requisitos definidos na secção 3.1, agrupados por critérios como eficiência, usabilidade, qualidade e segurança, após um período de utilização de cada nova versão do sistema, resultante das iterações da etapa de desenvolvimento, que lhes permitisse avaliar os resultados produzidos pelo STAAR.

De acordo com a metodologia descrita na secção 1.4., a escala de avaliação definida é:

- Não Atingido (N)
- Parcialmente Atingido (P)
- Largamente Atingido (L)
- Totalmente Atingido (F)

Tabela 10 - Avaliação do cumprimento dos requisitos por iteração DSRM

Critérios	Requisito	A solução deve…	1ª Iteração	2ª Iteração	3ª Iteração
Eficiência	1	ser capaz de transcrever automaticamente a fala em texto	P	L	F
	2	produzir transcrições com um texto legível e compreensível	L	L	F
	4	processar rapidamente os áudios assim que disponíveis	P	F	F
Usabilidade e Compatibilidade	3	permitir a edição e formatação do texto diretamente no Word	N	F	F
	5	ser compatível com equipamentos existentes	F	F	F
	6	permitir uso de hardware para controlo do áudio por transcrição	N	F	F
	7	identificar mudança de orador e associar texto transcrito	N	N	F
Flexibilidade	8	obter áudios de várias localizações	N	F	F
	9	transcrever áudios introduzidos manualmente	F	F	F
	12	ser expansível para outros contextos parlamentares	N	L	F
Especificidade Linguística	10	considerar as particularidades da língua portuguesa	F	F	F
	13	permitir a criação e personalização de dicionários	N	L	F
Qualidade	11	permitir avaliações e testes	F	F	F
	14	ter um WER abaixo de 15%	F	F	F
	18	lidar com ruídos de fundo	L	L	F
Segurança e Privacidade	15	processar os áudios na infraestrutura da AR	N	F	F
	16	permitir capacidade de trabalho remoto	F	F	F
	17	fornecer diferentes níveis de permissões de acesso	N	F	F

4.4. *Feedback* dos utilizadores

Uma implementação bem sucedida de qualquer sistema tecnológico não se baseia apenas na sua funcionalidade ou eficiência técnica, mas também na sua aceitação e utilidade percebida pelos utilizadores finais. A opinião dos utilizadores é muito importante para avaliar a eficácia de uma solução, identificar áreas de melhoria e garantir que o sistema responde às necessidades e expectativas.

Esta secção centra-se na recolha e análise do *feedback* dos utilizadores relativamente ao STAAR, de forma a fornecer informações valiosas sobre a sua experiência, satisfação e sugestões para futuras melhorias. Através de um inquérito procurou-se obter uma imagem holística da interação dos utilizadores com o sistema e do seu impacto no fluxo de trabalho parlamentar, relativamente à transcrição de debates parlamentares. A escolha de um inquérito como instrumento de recolha de dados foi motivada pela sua capacidade de alcançar um grande número de utilizadores de forma sistemática, permitindo a recolha de dados quantitativos e qualitativos.

O inquérito, cujo enunciado pode ser consultado no Anexo IV - Inquérito sobre o Sistema de Transcrição Automática (STAAR), é composto por duas secções e foi desenhado para abordar as principais áreas de interesse, desde a usabilidade e eficiência do sistema, até à sua integração no ambiente de trabalho parlamentar.

A primeira secção foca-se na caracterização do inquirido, para entender o perfil demográfico e profissional dos utilizadores através de aspetos como idade, serviço a que pertencem e anos de experiência em transcrição. Estes dados são fundamentais para contextualizar as respostas e perceber se existem padrões ou tendências específicas associadas a determinados grupos de utilizadores. A segunda secção centra-se na experiência direta com o Sistema de Transcrição Automático (STAAR), onde os utilizadores são convidados a partilhar as suas opiniões sobre a usabilidade, eficiência, precisão e outros aspetos relevantes do sistema. Através destas questões, procurou-se compreender o grau de satisfação dos utilizadores, identificar pontos fortes e também áreas de melhoria.

Para a resposta às diversas questões foi utilizada uma escala de classificação do tipo *Likert* [64] de 5 graus, com descrições verbais que contemplam extremos, como por exemplo “Sim, muito” e “Não de todo”. Esta escala permite que os inquiridos indiquem seu nível de discordância ou concordância em relação às questões colocadas, através uma escala ordinal de 1 a 5, o que facilita o tratamento dos dados e comparação de resultados. Esta abordagem possibilita a extração de informações quantitativas relativamente a perguntas estruturada de forma qualitativa.

Para a recolha de *feedback* dos utilizadores sobre o STAAR, optou-se por utilizar o Google Forms [65], uma ferramenta online que permite a criação de inquéritos personalizados de forma intuitiva e eficaz. A escolha desta plataforma *online* baseou-se na sua acessibilidade generalizada, flexibilidade na criação de questões, capacidade de análise integrada e garantia de confidencialidade nas respostas.

A 1 de novembro de 2023 foi elaborado e enviado um email com o *link* para os inquiridos, onde se explicava o propósito do inquérito, a importância do *feedback* para a melhoria contínua do STAAR e se garantia a confidencialidade das respostas.

Uma semana após o envio do inquérito por email, foram obtidas um total de 22 respostas, 19 relativas a funcionários da Divisão de Redação (DR) e 3 da Divisão de Apoio às Comissões (DAC), conforme ilustrado na Figura 38. Os resultados do inquérito podem ser consultados em maior detalhe no Anexo V - Resultados do inquérito ao STAAR. É de realçar que todos funcionários da AR dedicados à transcrição de debates parlamentares responderam ao inquérito, o que permite ter uma visão clara e global da perceção e experiência dos utilizadores na utilização do STAAR para as suas funções.

Em termos de distribuição etária, os inquiridos apresentam uma variedade significativa, entre os 30 e mais de 60 anos, que refletem diferentes fases da vida pessoal. Quanto à experiência profissional, observa-se uma distribuição equitativa entre funcionários com menos de 9 anos de experiência nas funções atuais, o que evidencia o esforço contínuo da AR na renovação dos seus quadros, e participantes com mais de 10 anos de experiência, o que demonstra também que a AR tem profissionais altamente experientes na área de transcrição de debates parlamentares.

Na Figura 24 é possível observar a distribuição dos inquiridos, em termos de idade e experiência profissional nas funções atuais relacionadas com transcrição.

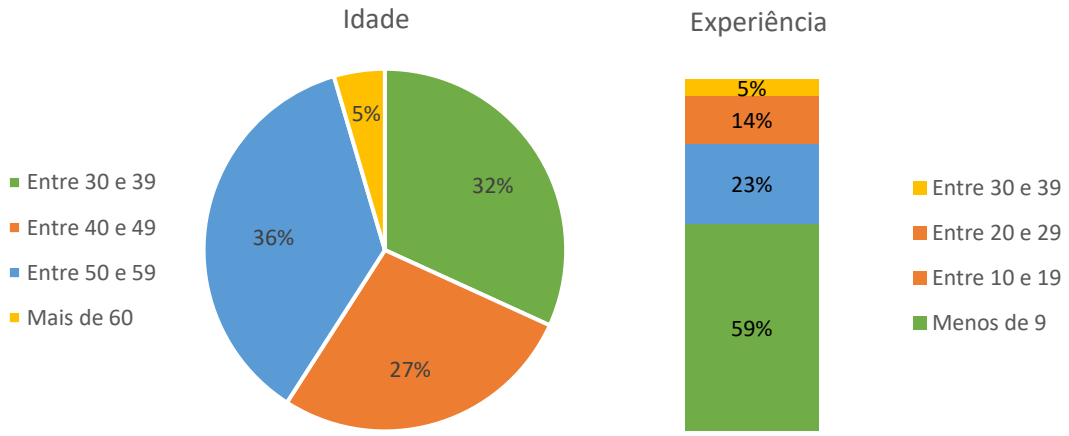


Figura 24 - Distribuição de inquiridos por idade e experiência

Relativamente à frequência de utilização do STAAR, como evidenciado na Figura 39, observa-se um uso intensivo do sistema, com a maior parte dos inquiridos a reportar recorrer ao STAAR de forma diária (45,5%) e semanal (40,9%).

As respostas obtidas foram estruturadas por forma a obter informação estatística, como a média, a mediana, valores mínimos e máximos bem como o desvio padrão das respostas. Com base nesta análise verificam-se os dados estatísticos que constam na Tabela 11.

Tabela 11 - Estatística descritiva das respostas ao questionário

Questão	Média	Mediana	Mínimo	Máximo	Desvio Padrão
Facilidade uso	4,6	5	4	5	0,5
Rapidez de transcrição	4,4	5	3	5	0,7
Precisão	3,7	4	3	5	0,6
Separação texto por orador	4,5	5	3	5	0,7
Formatação do texto	4,3	4	2	5	0,8
Responde às necessidades	4,3	4	4	5	0,5
Impacto positivo	4,6	5	4	5	0,5

O gráfico constante na Figura 25 apresenta uma visão abrangente das medianas das respostas obtidas no inquérito, o que permite uma perspetiva clara e comparativa sobre as diversas dimensões avaliadas pelos utilizadores do STAAR.

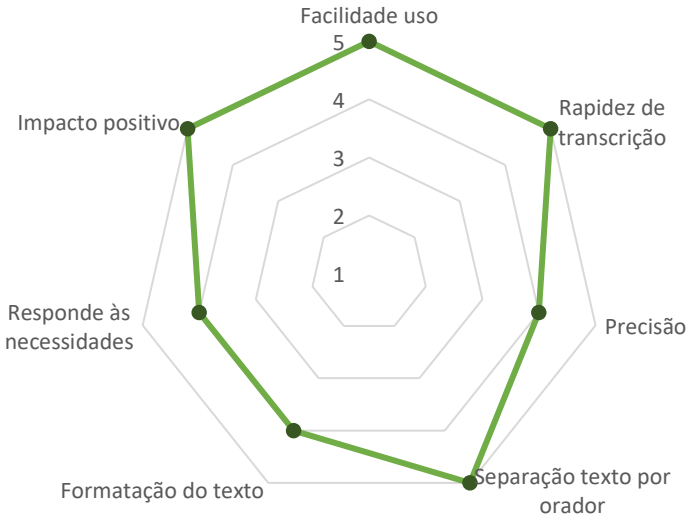


Figura 25 - Gráfico radar com mediana das respostas ao inquérito

Em geral, os desvios padrões mais baixos nas categorias de facilidade de uso, resposta às necessidades e impacto positivo sugerem uma perceção consistente e amplamente favorável do STAAR nestes aspetos. Por outro lado, a maior variação nas respostas sobre a formatação do texto e a rapidez de transcrição indica áreas onde as experiências dos utilizadores são mais diversas e onde podem ser necessárias melhorias ou ajustes.

A análise global das respostas indica uma receção bastante positiva do STAAR. A totalidade dos inquiridos classificou a facilidade de uso das transcrições (Figura 40) como "Fácil" (40,9%) ou "Muito fácil" (59,1%), e a rapidez com que as transcrições são disponibilizadas (Figura 41) foi predominantemente avaliada como "Muito rápida" (54,5%). Em termos de precisão (Figura 42), as transcrições foram maioritariamente consideradas "Alta" (54,5%) ou "Aceitável" (40,9%), o que sugere uma performance bastante satisfatória do sistema, mas com possibilidade de ser melhorada.

A funcionalidade de separação de texto por orador (Figura 43), possível através da diarização, foi bem recebida, com a esmagadora maioria dos inquiridos a considerá-la útil (90,9%). A formatação do texto e a identificação de correções automáticas (Figura 44) foram vistas como facilitadoras do trabalho dos utilizadores (86,4%), por melhorarem a experiência do processo de revisão e edição das transcrições automáticas.

No que se refere à adequação do STAAR às necessidades de transcrição dos áudios, todos os inquiridos consideraram que o sistema atende total (31,8%) ou parcialmente (68,2%) às suas necessidades (Figura 45). Verifica-se ainda que a implementação do STAAR foi vista como tendo um impacto positivo na função desempenhada pela totalidade dos inquiridos (Figura 46), com a maior expressão em “Muito” (59,1%), o que demonstra uma melhoria significativa em relação aos métodos anteriores de transcrição.

Entre as funcionalidades adicionais apresentadas aos inquiridos (Figura 47), destacam-se o desenvolvimento de um modelo linguístico próprio baseado em registos da AR para reduzir a taxa de erro nas transcrições (18%), a disponibilização numa só interface do áudio e do texto transcrito com pedaleira/atalhos para controlo do áudio (25%), e com maior expressão a identificação exata do nome/partido do orador no texto transcrito (27%). De destacar ainda a funcionalidade de produção de resumos/sumários que, apesar de no geral ter uma expressão baixa (7%), foi uma resposta escolhida por todos os inquiridos da área da DAC, sugerindo que deva ser igualmente levada em consideração. Todas estas sugestões apontam para áreas de melhoria e inovação que podem ser exploradas em futuras iterações do STAAR.

Face ao exposto, constata-se que o *feedback* dos utilizadores reflete uma elevada satisfação geral com o STAAR, que evidencia a sua relevância na transformação e otimização do processo de transcrição no contexto parlamentar. Através da sua rapidez, precisão e funcionalidades inovadoras, o STAAR não só dá resposta às necessidades dos seus utilizadores, como também se afirma como uma ferramenta valiosa na modernização e eficiência da documentação parlamentar.

4.5. Dificuldades encontradas

Ao longo do desenvolvimento e implementação do Sistema de Transcrição Automático da Assembleia da República (STAAR), foram encontradas diversas dificuldades que exigiram uma abordagem adaptativa e soluções inovadoras, mas que se enquadrassem no contexto da Assembleia da República.

Em relação aos desafios técnicos, a utilização do modelo de linguagem Whisper, apesar de inovador, apresentou desafios significativos. Dada a sua novidade, havia uma falta de documentação detalhada e casos de implementação o que tornou a integração e otimização do Whisper para o contexto parlamentar um desafio técnico considerável. A ausência de *feedback* e experiências de outros utilizadores na comunidade científica/profissional também limitou a capacidade de antecipar e resolver problemas específicos relacionados com este modelo. Para superar estes desafios, foi necessário bastante tempo por parte dos especialistas em tecnologia da informação e realização de testes contínuos até chegar a uma solução.

A nível operacional, a transição de um sistema manual para um sistema automático é sempre desafiadora. A implementação do STAAR exigiu a introdução de novas formas de trabalhar, algo a que os seres humanos, por natureza, tendem a resistir. Durante a fase de desenho, o Grupo de Trabalho da Transcrição Automática (GT-TA) frequentemente solicitava que o STAAR replicasse os métodos de trabalho existentes, em vez de adotar abordagens mais inovadoras. Embora a inovação fosse o objetivo, houve momentos em que foi necessário ceder a esses pedidos para garantir a aceitação e a adoção do sistema pelos utilizadores. A formação contínua dos utilizadores e a disponibilização de suporte técnico foram essenciais para superar estes desafios.

Quanto aos desafios contextuais, o ambiente parlamentar tem as suas peculiaridades. O jargão específico usado neste contexto e o formato rígido do Diário da Assembleia da República (DAR) apresentaram desafios únicos. Foi necessário customizar e formatar o sistema para reconhecer e transcrever corretamente algum palavreado específico do parlamento. Além disso, a necessidade de garantir que o texto transcrito se alinhasse com as normas e formatos estabelecidos do DAR exigiu várias otimizações e ajustes no sistema. A colaboração com a equipa de transcrição da Assembleia da República foi chave para adaptar o sistema às necessidades específicas do ambiente parlamentar.

A implementação do STAAR foi um processo repleto de desafios, mas com a combinação certa de *expertise* técnico e compreensão profunda do contexto parlamentar do GT-TA, foi possível superar os obstáculos e criar uma solução robusta e eficaz no processo de transcrição de debates parlamentares.

CAPÍTULO 5.

Conclusões e Trabalho Futuro

5.1. Conclusões

O desenvolvimento do Sistema de Transcrição Automática para a Assembleia da República (STAAR) representa um avanço significativo no trabalho de transcrição dos debates parlamentares, alinhando-se estreitamente com os objetivos estabelecidos no início deste trabalho de projeto de mestrado.

O primeiro objetivo, focado na identificação e análise de tecnologias de reconhecimento automático de fala (*speech-to-text*) e sua aplicabilidade na transcrição de debates parlamentares, foi atingido através de uma investigação aprofundada do estado da arte em transcrição automática, não só através de revisão de literatura, mas também com base na experiência de outros Parlamentos. Esta análise detalhada permitiu a seleção do modelo de linguagem Whisper, que se destacou pela sua facilidade de uso e adequação às especificidades linguísticas e acústicas do ambiente parlamentar.

Através da implementação de tecnologias avançadas de Inteligência Artificial, como o modelo de linguagem Whisper, foi possível automatizar e otimizar um processo que, tradicionalmente, dependia de esforços manuais intensivos e que envolviam um desgaste físico acentuado. A arquitetura modular do STAAR, composta por etapas de recolha de áudios, processamento de áudio, tratamento de texto e armazenamento, demonstrou ser eficaz para responder às necessidades específicas do ambiente parlamentar em termos de transcrição de áudio para texto. Além disso, a capacidade de adaptar-se às particularidades da linguagem parlamentar e ao jargão específico, bem como a integração com sistemas existentes, destacou a flexibilidade e robustez do sistema e permitiu atingir em pleno o segundo objetivo deste trabalho.

O terceiro e último objetivo, a implementação de tecnologias de reconhecimento de mudança de orador para melhorar a transcrição de debates parlamentares, foi alcançado com sucesso através de métodos de diarização no STAAR. Esta funcionalidade enriqueceu as transcrições e facilitou a leitura e a revisão através da clara segmentação visual do texto por orador, o que melhorou a precisão e a utilidade das transcrições automáticas.

Os resultados globais obtidos com o STAAR superaram as expectativas não só por ser extremamente rápido a produzir as transcrições, mas também devido ao facto da taxa de erro (WER) destas ser muito inferior ao que inicialmente foi considerado como aceitável.

A incorporação de funcionalidades avançadas, como a correção automática de texto via dicionários especializados, a identificação de alternância entre oradores e a formatação adequada do

texto, culminou em documentos que se assemelham, em grande medida, aos anteriormente realizados de forma manual no contexto da Assembleia da República (AR).

Para a equipa de transcrição da AR, este sistema representou uma transformação profunda na forma como o seu trabalho é realizado, ao eliminar por completo a necessidade de transcrição manual dos debates parlamentares, ainda que a tarefa de revisão se tenha tornado mais exigente, fruto da necessidade da identificação e correção de eventuais erros de transcrição.

Com isto, a implementação do STAAR não só melhorou a eficiência das transcrições, reduziu o tempo necessário para a produção da primeira versão interna do Diário da Assembleia da República, mas também proporcionou uma economia significativa de recursos, ao permitir que os profissionais de transcrição se concentrem em tarefas mais complexas e menos rotineiras.

Adicionalmente, devido aos resultados bastante positivos alcançados pelo STAAR, surgiu rapidamente a solicitação para que o sistema fosse utilizado na transcrição de reuniões de comissões parlamentares, tarefa esta que anteriormente não era realizada devido à limitação de recursos humanos, ampliando assim a abrangência e o detalhe na documentação das atividades parlamentares.

Face ao exposto, conclui-se que o Sistema de Transcrição Automática (STAAR) não só deu resposta aos problemas e requisitos identificados, ao cumprir todos os objetivos deste trabalho de projeto, como também adicionou valor significativo à Assembleia da República, posicionando-a na vanguarda da inovação tecnológica no contexto da transcrição de debates parlamentares.

5.2. Trabalho futuro

Apesar dos avanços significativos alcançados com o STAAR no âmbito da transcrição de debates parlamentares, a natureza dinâmica da tecnologia e as necessidades em constante evolução da Assembleia da República sugerem que há sempre espaço para melhorias e expansões.

Ao refletir sobre o trabalho realizado e ao antecipar as necessidades futuras, é possível identificar várias oportunidades de melhoria que não só visam otimizar o sistema existente, mas também expandir sua aplicabilidade e garantir sua relevância contínua, entre as quais se destacam:

- Otimização do modelo de linguagem - Com o avanço contínuo da tecnologia de IA, há espaço para aperfeiçoar cada vez mais o modelo de linguagem utilizado, tornando-o mais preciso e adaptado às nuances dos debates parlamentar;
- Criação de *corpus* de oradores - Esta inovação permitiria ao STAAR identificar com precisão o autor de cada intervenção parlamentar. Ao reconhecer e atribuir automaticamente o nome e o partido do orador à transcrição, seria possível eliminar uma tarefa manual que atualmente

é realizada pela equipa de transcrição, otimizando ainda mais o processo, entre outras aplicações possíveis;

- Redução do tempo global de transcrição - Visto a etapa de diarização representar cerca de 85% do tempo necessário para a transcrição de um áudio, a sua otimização poderia reduzir significativamente a duração global do processo, de forma a aumentar a eficiência do sistema e acelerar a disponibilização das transcrições para uso parlamentar;
- Substituição contextual probabilística - Uma evolução significativa para o STAAR seria a introdução de um mecanismo de substituição de texto baseado em aproximação que iria além da simples correspondência de palavras ou expressões, incorporando uma análise contextual do texto de transcrição. Através de abordagens probabilísticas, o sistema poderia identificar e corrigir erros com base em limiares de correspondência, de forma a garantir correções mais precisas e adaptadas ao contexto em que as palavras ou expressões aparecem;
- Adoção de um formato de transcrição estruturado - Ao migrar de um formato tradicional como o docx para um formato mais estruturado e interoperável seria possível realizar consultas segmentadas por orador, tema ou qualquer outro critério relevante o que potenciaria um tratamento mais ágil e personalizado das transcrições. Esta mudança não só beneficiaria os Deputados, mas também facilitaria o acesso e a compreensão do público em geral sobre os debates parlamentares;
- Criação de resumos de forma automática - A implementação de algoritmos avançados de sumarização automática poderia transformar extensos registos textuais em sínteses claras e relevantes, e potenciar a eficiência do trabalho por exemplo de Comissões Parlamentares;
- Integração com outras plataformas - O STAAR pode ser integrado com outras plataformas digitais da Assembleia da República, e permitir uma disseminação mais ampla e acessível das transcrições.

Referências Bibliográficas

[1] "História do jornal oficial do Parlamento". Acesso em: 20 de maio de 2023. [Online]. Disponível em: https://www.parlamento.pt:443/DAR/Paginas/HistoriaJornal.aspx

[2] "Debates Parlamentares". Acesso em: 20 de maio de 2023. [Online]. Disponível em: https://debates.parlamento.pt/

[3] X. Huang, J. Baker, e R. Reddy, "A historical perspective of speech recognition", *Commun. ACM*, vol. 57, nº 1, p. 94–103, jan. 2014, doi: 10.1145/2500887.

[4] A. Stolcke, "SRILM - an extensible language modeling toolkit", em *7th International Conference on Spoken Language Processing (ICSLP 2002)*, ISCA, set. 2002, p. 901–904. doi: 10.21437/ICSLP.2002-303.

[5] A. Kumar, S. Verma, e H. Mangla, "A Survey of Deep Learning Techniques in Speech Recognition", em *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, out. 2018, p. 179–185. doi: 10.1109/ICACCCN.2018.8748399.

[6] K. Peffers, T. Tuunanen, M. A. Rothenberger, e S. Chatterjee, "A design science research methodology for information systems research", *Journal of Management Information Systems*, vol. 24, nº 3, p. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.

[7] 14:00-17:00, "ISO/IEC 15504-2:2003", ISO. Acesso em: 23 de outubro de 2023. [Online]. Disponível em: https://www.iso.org/standard/37458.html

[8] K. El Emam, "The internal consistency of the ISO/IEC 15504 software process capability scale", em *Proceedings Fifth International Software Metrics Symposium. Metrics (Cat. No.98TB100262)*, nov. 1998, p. 72–81. doi: 10.1109/METRIC.1998.731228.

[9] "WER | Calculate the Word Error Rate with our Tool", Amberscript. Acesso em: 21 de outubro de 2023. [Online]. Disponível em: https://www.amberscript.com/en/wer-tool/

[10] L. Besacier, E. Barnard, A. Karpov, e T. Schultz, "Automatic speech recognition for under-resourced languages: A survey", *Speech Communication*, vol. 56, nº 1, p. 85–100, 2014, doi: 10.1016/j.specom.2013.07.008.

[11] K. H. Davis, R. Biddulph, e S. Balashek, "Automatic Recognition of Spoken Digits", *Journal of the Acoustical Society of America*, vol. 24, nº 6, p. 637–642, 1952, doi: 10.1121/1.1906946.

[12] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, vol. 77, nº 2, p. 257–286, 1989, doi: 10.1109/5.18626.

[13] A. Virkkunen, A. Rouhe, N. Phan, e M. Kurimo, "Finnish parliament ASR corpus: Analysis, benchmarks and statistics", *Lang. Resour. Eval.*, 2023, doi: 10.1007/s10579-023-09650-7.

[14] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", *IEEE Signal Processing Magazine*, vol. 29, nº 6, p. 82–97, 2012, doi: 10.1109/MSP.2012.2205597.

[15] G. E. Dahl, D. Yu, L. Deng, e A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition", *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, nº 1, p. 30–42, 2012, doi: 10.1109/TASL.2011.2134090.

[16] A. Graves, A.-R. Mohamed, e G. Hinton, "Speech recognition with deep recurrent neural networks", apresentado em ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2013, p. 6645–6649. doi: 10.1109/ICASSP.2013.6638947.

[17] A. Hannun *et al.*, “Deep Speech: Scaling up end-to-end speech recognition”. arXiv, 19 de dezembro de 2014. doi: 10.48550/arXiv.1412.5567.

[18] S. Karita *et al.*, “A Comparative Study on Transformer vs RNN in Speech Applications”, apresentado em 2019 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2019 - Proceedings, 2019, p. 449–456. doi: 10.1109/ASRU46091.2019.9003750.

[19] A. Gulati *et al.*, “Conformer: Convolution-augmented transformer for speech recognition”, apresentado em Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2020, p. 5036–5040. doi: 10.21437/Interspeech.2020-3015.

[20] A. Baevski, H. Zhou, A. Mohamed, e M. Auli, “wav2vec 2.0: A framework for self-supervised learning of speech representations”, apresentado em Advances in Neural Information Processing Systems, 2020.

[21] Y.-A. Chung, W.-H. Weng, S. Tong, e J. Glass, “Towards Unsupervised Speech-to-text Translation”, apresentado em ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2019, p. 7170–7174. doi: 10.1109/ICASSP.2019.8683550.

[22] “List of countries and dependencies by population”, *Wikipedia*. 2 de junho de 2023. Acesso em: 3 de junho de 2023. [Online]. Disponível em: https://en.wikipedia.org/w/index.php?title=List_of_countries_and_dependencies_by_population&oldid=1158206585

[23] A. Hämäläinen *et al.*, “Automatically recognising European Portuguese children’s speech: Pronunciation patterns revealed by an analysis of ASR errors”, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8775, p. 1–11, 2014, doi: 10.1007/978-3-319-09761-9.

[24] T. Pellegrini *et al.*, “A corpus-based study of elderly and young speakers of european portuguese: Acoustic correlates and their impact on speech recognition performance”, apresentado em Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2013, p. 852–856.

[25] T. Aguiar de Lima e M. Da Costa-Abreu, “A survey on automatic speech recognition systems for Portuguese language and its variations”, *Computer Speech and Language*, vol. 62, 2020, doi: 10.1016/j.csl.2019.101055.

[26] “ECPRD | European Center for Parliamentary Research and Documentation”. Acesso em: 3 de junho de 2023. [Online]. Disponível em: https://ecprd.secure.europarl.europa.eu/ecprd/public/page/about

[27] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”, *BMJ*, vol. 372, p. n71, mar. 2021, doi: 10.1136/bmj.n71.

[28] “Scopus - Document search”. Acesso em: 21 de maio de 2023. [Online]. Disponível em: https://www.scopus.com/search/form.uri?display=basic#basic

[29] “IEEE Xplore - Document search”. Acesso em: 21 de maio de 2023. [Online]. Disponível em: https://ieeexplore.ieee.org/Xplore/home.jsp

[30] “Web of Science Core Collection - Document search”. Acesso em: 21 de maio de 2023. [Online]. Disponível em: https://www.webofscience.com/wos/woscc/basic-search

[31] “Zotero | Your personal research assistant”. Acesso em: 6 de maio de 2023. [Online]. Disponível em: https://www.zotero.org/start

[32] “Software de Folha de Cálculo do Microsoft Excel | Microsoft 365”. Acesso em: 21 de maio de 2023. [Online]. Disponível em: https://www.microsoft.com/pt-pt/microsoft-365/excel

[33] W. X. Zhao *et al.*, “A Survey of Large Language Models”. arXiv, 7 de maio de 2023. Acesso em: 28 de maio de 2023. [Online]. Disponível em: http://arxiv.org/abs/2303.18223

[34] H. Vos e S. Verberne, *Political corpus creation through automatic speech recognition on EU debates*. 2023.

[35] C. V. G. Diáz-Munió *et al.*, “Europarl-ASR: A large corpus of parliamentary debates for streaming ASR benchmarking and speech data filtering/verbatimization”, em *Proc. Annu. Conf. Int. Speech. Commun. Assoc., INTERSPEECH*, International Speech Communication Association, 2021, p. 4371–4375. doi: 10.21437/Interspeech.2021-1905.

[36] T. Alumaë, O. Tilk, e A. Ullah, *Advanced rich transcription system for Estonian speech*, vol. 307. em Frontiers in Artificial Intelligence and Applications, vol. 307. 2018, p. 8. doi: 10.3233/978-1-61499-912-6-1.

[37] T. Kawahara, “Automatic meeting transcription system for the Japanese parliament (diet)”, em *Proc. - Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf., APSIPA ASC*, Institute of Electrical and Electronics Engineers Inc., 2018, p. 1006–1010. doi: 10.1109/APSIPA.2017.8282177.

[38] A. Mansikkaniemi, P. Smit, e M. Kurimo, “Automatic construction of the Finnish parliament speech corpus”, em *Proc. Annu. Conf. Int. Speech. Commun. Assoc., INTERSPEECH*, Lacerda F., Strombergsson S., Wlodarczak M., Heldner M., Gustafson J., e House D., Orgs., International Speech Communication Association, 2017, p. 3762–3766. doi: 10.21437/Interspeech.2017-1115.

[39] F. De Wet, J. Badenhorst, e T. Modipa, “Developing Speech Resources from Parliamentary Data for South African English”, apresentado em Procedia Computer Science, 2016, p. 45–52. doi: 10.1016/j.procs.2016.04.028.

[40] P. Campr, M. Kunešová, J. Vaněk, J. Čech, e J. Psutka, “Audio-video speaker diarization for unsupervised speaker and face model creation”, em *Lect. Notes Comput. Sci.*, Springer Verlag, 2014, p. 465–472. doi: 10.1007/978-3-319-10816-2_56.

[41] “HTK Speech Recognition Toolkit”. Acesso em: 29 de outubro de 2023. [Online]. Disponível em: https://htk.eng.cam.ac.uk/

[42] “Kaldi: Kaldi”. Acesso em: 29 de outubro de 2023. [Online]. Disponível em: https://kaldi-asr.org/doc/index.html

[43] “Multilinguismo no Parlamento Europeu”, Multilinguismo. Acesso em: 29 de outubro de 2023. [Online]. Disponível em: https://www.europarl.europa.eu/about-parliament/pt/organisation-and-rules/multilingualism

[44] “LIBE | Comissões | Parlamento Europeu”. Acesso em: 29 de outubro de 2023. [Online]. Disponível em: https://www.europarl.europa.eu/committees/pt/libe/home/highlights

[45] J. Devlin, M.-W. Chang, K. Lee, e K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. arXiv, 24 de maio de 2019. doi: 10.48550/arXiv.1810.04805.

[46] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach”. arXiv, 26 de julho de 2019. doi: 10.48550/arXiv.1907.11692.

[47] R. Ma, M. J. F. Gales, K. M. Knill, e M. Qian, “N-best T5: Robust ASR Error Correction using Multiple Input Hypotheses and Constrained Decoding Space”, em *INTERSPEECH 2023*, ago. 2023, p. 3267–3271. doi: 10.21437/Interspeech.2023-1616.

[48] OpenAI, “GPT-4 Technical Report”. arXiv, 27 de março de 2023. doi: 10.48550/arXiv.2303.08774.

[49] “LIUM SpkDiarization”. Acesso em: 29 de outubro de 2023. [Online]. Disponível em: https://projets-lium.univ-lemans.fr/spkdiarization/

[50] “Introducing Whisper”. Acesso em: 25 de setembro de 2023. [Online]. Disponível em: https://openai.com/research/whisper

[51] “GPT-4”. Acesso em: 5 de novembro de 2023. [Online]. Disponível em: https://openai.com/gpt-4

[52] “DALL·E 3”. Acesso em: 5 de novembro de 2023. [Online]. Disponível em: https://openai.com/dall-e-3

[53] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, e I. Sutskever, “Robust Speech Recognition via Large-Scale Weak Supervision”, 2022, doi: 10.48550/ARXIV.2212.04356.

[54] A. Vaswani *et al.*, “Attention is all you need”, apresentado em Advances in Neural Information Processing Systems, 2017, p. 5999–6009. [Online]. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85043317328&partnerID=40&md5=3e5a5c2b862c8979ffea845bb707b3c3

[55] “whisper/approach.png at main · openai/whisper”, GitHub. Acesso em: 27 de setembro de 2023. [Online]. Disponível em: https://github.com/openai/whisper/blob/main/approach.png

[56] “CoVoST: A Large-Scale Multilingual Speech-To-Text Translation Corpus”. Meta Research, 15 de setembro de 2023. Acesso em: 25 de setembro de 2023. [Online]. Disponível em: https://github.com/facebookresearch/covost

[57] M. Bain, J. Huh, T. Han, e A. Zisserman, “WhisperX: Time-Accurate Speech Transcription of Long-Form Audio”. arXiv, 11 de julho de 2023. doi: 10.48550/arXiv.2303.00747.

[58] M. Bain, “WhisperX”. 26 de setembro de 2023. Acesso em: 26 de setembro de 2023. [Online]. Disponível em: https://github.com/m-bain/whisperX

[59] H. Bredin *et al.*, “pyannote.audio: neural building blocks for speaker diarization”. arXiv, 4 de novembro de 2019. doi: 10.48550/arXiv.1911.01255.

[60] “The official home of the Python Programming LanguageWebsite Python.org”, Python.org. Acesso em: 25 de setembro de 2023. [Online]. Disponível em: https://www.python.org/

[61] “BPMN Specification - Business Process Model and Notation”. Acesso em: 2 de julho de 2023. [Online]. Disponível em: https://www.bpmn.org/

[62] “Google Colaboratory”. Acesso em: 19 de outubro de 2023. [Online]. Disponível em: https://colab.research.google.com/

[63] “Acordo Ortográfico - Portal da Língua Portuguesa”. Acesso em: 24 de outubro de 2023. [Online]. Disponível em: http://www.portaldalinguaportuguesa.org/acordo.php?action=acordo&version=1990b

[64] R. Likert, S. Roslow, e G. Murphy, “A Simple and Reliable Method of Scoring the Thurstone Attitude Scales”, *Journal of Social Psychology*, vol. 5, nº 2, p. 228–238, 1934, doi: 10.1080/00224545.1934.9919450.

[65] “Google Forms: criador de formulários online | Google Workspace”. Acesso em: 13 de junho de 2023. [Online]. Disponível em: https://www.facebook.com/GoogleDocs/

Anexos

Anexo I - Código das funções python desenvolvidas

Ao longo do desenvolvimento do STAAR, diversas funções foram desenvolvidas e aperfeiçoadas, com o objetivo de assegurar a eficácia do sistema. Este anexo detalha as funções programadas em Python [60], cada uma cuidadosamente elaborada para responder às necessidades específicas identificadas no capítulo 3, Desenho e Desenvolvimento.

Este anexo não só serve como uma referência técnica para os interessados nos aspetos de programação e implementação do STAAR, mas também evidencia a complexidade e a inovação das soluções desenvolvidas para superar os desafios inerentes à transcrição automática no contexto parlamentar.

```
    # Procurar ficheiros mp3 criados dentro do "cutoff_time" e que ainda não tenham sido transcritos
    for root, dirs, files in os.walk(source_folder):
        for file in files:
            file_path = os.path.join(root, file)
            if file_path.endswith(".mp3") and os.path.getctime(file_path) > cutoff_time.timestamp():
                try:
                    transcribe_file(file_path)
                except Exception as e:
                    logging.info(f"Erro: {e}")
```

Figura 26 - Função de procura de ficheiros por transcrever

```
def whisper_transcribe (audio_file, output_file):
    # Variáveis
    global transcription_duration, diarization_duration, alignment_duration

    # 1. Transcrição com Whisper (batched)
    transcription_start_time = get_current_time()
    logging.info(f"Transcribing {(audio_file)}...")
    audio = whisper.load_audio(audio_file)
    result = model.transcribe(audio, batch_size=whisper_batch_size)
    transcription_end_time = get_current_time()
```

Figura 27 - Função de Transcrição

```
    # 2. Alinhamento do output do Whisper, através do WhisperX
    alignment_start_time = get_current_time()
    logging.info(f"Aligning {(audio_file)}...")
    model_a, metadata = whisperx.load_align_model(language_code=whisperx_language, device=
    whisperx_device)
    #model_a, metadata = whisperx.load_align_model(language_code=result["language"],
    device=whisperx_device)
    result = whisperx.align(result["segments"], model_a, metadata, audio, whisperx_device,
    return_char_alignments=False)
    alignment_end_time = get_current_time()
```

Figura 28 - Função de Alinhamento

```
    # 3. Diarização com WhisperX
    diarization_start_time = get_current_time()
    logging.info(f"Diarizing {(audio_file)}...")
    diarize_model = whisperx.DiarizationPipeline(use_auth_token=whisperx_hf_token, device=
    whisperx_device)
    diarize_segments = diarize_model(audio)
    result = whisperx.assign_word_speakers(diarize_segments, result)
    diarization_end_time = get_current_time()

    # Cálculo de duração dos módulos
    transcription_duration = int((transcription_end_time - transcription_start_time).total_seconds())
    alignment_duration = int((alignment_end_time - alignment_start_time).total_seconds())
    diarization_duration = int((diarization_end_time - diarization_start_time).total_seconds())
```

Figura 29 - Função de Diarização

```
def parse_transcript(input_filename, output_filename):
    with open(input_filename, 'r', encoding="utf-8") as file:
        lines = file.readlines()
    output = []
    current_speaker = None
    current_text = ""
    for line in lines:
        if not line.startswith('['):  # saltar linhas sem orador
            continue
        speaker_id = line.split(']:')[0][1:]
        text = line.split(']:')[1].strip()
        speaker_id_parts = speaker_id.split('_')
        speaker_id_number = speaker_id_parts[-1]
        if current_speaker is None:
            current_speaker = speaker_id
            current_text = f"ORADOR_{speaker_id_number}: — {text}"
        elif current_speaker == speaker_id:
            current_text += " " + text
        else:
            output.append(current_text.strip())
            current_speaker = speaker_id
            current_text = f"ORADOR_{speaker_id_number}: — {text}"
    # Adicionar a última parte do texto
    if current_text.strip():  # verificar se está vazia
        output.append(current_text.strip())
```

Figura 30 - Função de Remoção Ids de orador e timestamps

```
def remove_phrases(input_file, phrases_file, output_file):
    # Abre o ficheiro de frases a remover em minúsculas e o ficheiro de input
    with open(phrases_file, 'r', encoding="utf-8") as phrases_file:
        phrases = [line.strip().lower() for line in phrases_file.readlines()]
    with open(input_file, 'r', encoding="utf-8") as input_file:
        # Lê todas as linhas do ficheiro de input
        input_lines = input_file.readlines()
    # Abre ficheiro de output
    with open(output_file, 'w', encoding="utf-8") as output_file:
        for line in input_lines:
            line = line.strip()
            # Compara linhas input com linhas frases e remove
            line_lower = line.lower()
            if not any(line_lower.startswith(phrase) for phrase in phrases):
                output_file.write(line + '\n')
```

Figura 31 - Função de Remoção de frases

```python
def count_words_in_file(file):
    with open(file, 'r', encoding="utf-8") as file:
        text = file.read()
        words = text.split()
        return len(words)
```

Figura 32 - Função de Contagem de palavras

```python
def replace_words_in_text(input, words, output, case_sensitive=False):
    # Lê texto de ficheiro input
    with open(input, 'r', encoding="utf-8") as txt_file:
        text = txt_file.read()
    # Lê as palavras do dicionário fornecido
    replacement_map = {}
    with open(words, 'r', encoding="utf-8") as f:
        for line in f:
            line = line.strip()
            if line:
                word_to_search, word_to_replace = re.split(r'\s*,\s*', line, maxsplit=1)
                if not case_sensitive:
                    word_to_search = word_to_search.lower()
                replacement_map[word_to_search] = word_to_replace
    # Rotina de substituição de palavras
    def replace_word(match):
        matched_word = match.group()
        key = matched_word.lower() if not case_sensitive else matched_word
        replacement = replacement_map.get(key, None)

        if replacement is not None and matched_word != replacement:
            return f'***{replacement}***'
        else:
            return matched_word
    # Substituição case insensitive
    if not case_sensitive:
        for word_to_search in replacement_map.keys():
            escaped_word_to_search = re.escape(word_to_search)
            pattern = re.compile(r"(?i)\b(?!<\*\*)" + escaped_word_to_search + r"(?!\*\*)(?=\W|$)")
            text = pattern.sub(replace_word, text)
    # Substituição case sensitive
    else:
        for word_to_search in replacement_map.keys():
            escaped_word_to_search = re.escape(word_to_search)
            pattern = re.compile(r"\b(?!<\*\*)" + escaped_word_to_search + r"(?!\*\*)(?=\W|$)")
            text = pattern.sub(replace_word, text)
```

Figura 33 - Função de Substituição de texto

```python
def txt_2_docx(input_file, output_file):
    # Abre o ficheiro input e lê o seu conteúdo
    with open(input_file, 'r', encoding="utf-8") as infile:
        content = infile.read()
    # Criar documento docx
    doc = docx.Document()
    # Divide o conteúdo em parágrafos com base no caracter "."
    paragraphs = content.split('\n')
    # Adiciona os parágrafos ao docx
    for paragraph_text in paragraphs:
        p = doc.add_paragraph()
        p.add_run(paragraph_text.strip())
```

Figura 34 - Função de Conversão para docx

```python
def color_and_remove_markers(input_file, output_file):
    # Abre ficheiro input e cria ficheiro output
    doc = Document(input_file)
    output_doc = Document()
    # Define o padrão de pesquisa, texto prefixado e sufixado por ***
    pattern = r'\*\*\*(.*?)\*\*\*'
    # Iteração por cada parágrafo
    for para in doc.paragraphs:
        new_para = output_doc.add_paragraph()
        while True:
            match = re.search(pattern, para.text)
            if match:
                # Obter o texto sem ***
                text_to_color = match.group(1)
                text_before = para.text[:match.start()]
                if text_before:
                    run = new_para.add_run(text_before)
                    font = run.font
                    font.color.rgb = RGBColor(0, 0, 0)  # Preto
                # Mudar texto entre *** para azul
                run = new_para.add_run(text_to_color)
                font = run.font
                font.color.rgb = RGBColor(0, 0, 255)  # Azul
                para.text = para.text[match.end():]
            else:
                break

        # Adicionar texto restante a preto
        if para.text:
            run = new_para.add_run(para.text)
            font = run.font
            font.color.rgb = RGBColor(0, 0, 0)  # Preto
```

Figura 35 - Função de Identificação de substituições

```python
def format_text(input, output, lspace_after, f_name, f_size, alignment, p_ident, p_lstline_ident,
l_spacing):
    # Abrid ficheiro input
    doc = Document(input)
    # Variáveis alinhamento Word
    if alignment.lower() == 'center':
        alignment_enum = WD_ALIGN_PARAGRAPH.CENTER
    elif alignment.lower() == 'right':
        alignment_enum = WD_ALIGN_PARAGRAPH.RIGHT
    elif alignment.lower() == 'justify':
        alignment_enum = WD_ALIGN_PARAGRAPH.JUSTIFY
    else:
        alignment_enum = WD_ALIGN_PARAGRAPH.LEFT
    # Iteração por parágrafo
    for para in doc.paragraphs:
        para.paragraph_format.space_after = Pt(lspace_after)
        para.alignment = alignment_enum  # mudar alinhamento
        # Muda identação de primeiro parágrafo
        para.paragraph_format.left_indent = Cm(p_ident)
        para.paragraph_format.first_line_indent = Cm(p_lstline_ident)
        # Define letra e tamanho
        for run in para.runs:
            run.font.name = f_name
            run.font.size = Pt(f_size)
        # Muda o espaçamento entre linhas
        para.paragraph_format.line_spacing = l_spacing
```

Figura 36 - Função de Formatação de documento

```
def write_to_dabatase(audio_file_name, audio_file_path, transcription_file_name,
transcription_file_path, audio_creation_date, process_start_time, process_end_time, process_duration,
 audiofile_duration, transcription_duration, alignment_duration, diarization_duration, word_count,
host_name, engine, org):
    # Liga à base de dados
    conn_str = f'DRIVER=SQL Server;SERVER={db_server};DATABASE={db_database};UID={db_username};PWD={
    db_password};'
    conn = pyodbc.connect(conn_str)
    # Cria cursos para executar comandos SQL
    cursor = conn.cursor()
    # Cria tabela se não existir
    create_table_if_not_exists(cursor, db_table)
    # Escreve dados na base de dados
    insert_sql = f"""
        INSERT INTO {db_table} (
            audio_file_name, audio_file_path, transcription_file_name, transcription_file_path,
            audio_creation_date, process_start_time, process_end_time, process_duration,
            audiofile_duration,
            transcription_duration, alignment_duration, diarization_duration, word_count, host_name,
            engine, org
        )
        VALUES (?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?)
    """
    cursor.execute(insert_sql, (
        audio_file_name, audio_file_path, transcription_file_name, transcription_file_path,
        audio_creation_date, process_start_time, process_end_time, process_duration,
        audiofile_duration,
        transcription_duration, alignment_duration, diarization_duration, word_count, host_name,
        engine, org
    ))
```

Figura 37 - Função de Gravação de informação em base de dados

Anexo II - Análise comparativa de transcrições

A Tabela 12, realizada no âmbito do Grupo de Trabalho da Transcrição Automática (GT-TA) durante a segunda iteração DSRM, mostra as diferenças entre uma transcrição manual de um discurso proferido pelo Presidente da Assembleia da República durante uma sessão solene, realizada pela equipa de transcrição da AR, com a gerada automaticamente pelo STAAR, sem recorrer a qualquer dicionário de substituição de palavras/termos. Do lado direito, Transcrição STAAR, são identificadas a vermelho as diferenças encontradas.

Tabela 12 - Análise comparativa de transcrições

Transcrição manual	Transcrição STAAR
SENHOR PRESIDENTE DA REPÚBLICA PORTUGUESA SENHOR PRESIDENTE DA REPÚBLICA DA UCRÂNIA SENHOR PRIMEIRO-MINISTRO E DEMAIS MEMBROS DO GOVERNO SENHORAS E SENHORES DEPUTADOS SENHOR EMBAIXADOR DA UCRÂNIA EM PORTUGAL ILUSTRES CONVIDADOS MINHAS SENHORAS E MEUS SENHORES PORTUGAL CONDENOU DESDE O PRIMEIRO MOMENTO COM FIRME DETERMINAÇÃO A AGRESSÃO MILITAR DA FEDERAÇÃO RUSSA CONTRA A UCRÂNIA MENOS DE TRÊS HORAS APÓS O SEU INÍCIO NA NOITE DE 23 PARA 24 DE FEVEREIRO O GOVERNO PORTUGUÊS REPROVOU-A PUBLICAMENTE NO DIA 24 TODOS OS ÓRGÃOS POLÍTICOS DE SOBERANIA O PRESIDENTE DA REPÚBLICA A ASSEMBLEIA DA REPÚBLICA E O GOVERNO CONDENAVAM EM UNÍSSONO O AGRESSOR E EXPRIMIAM SOLIDARIEDADE E APOIO AO AGREDIDO FIZERAM-NO ENTÃO E TÊM REITERADAMENTE FEITO SEM QUALQUER HESITAÇÃO NEM AMBIGUIDADE PARA PORTUGAL O AGRESSOR É A FEDERAÇÃO RUSSA E O AGREDIDO É A UCRÂNIA O ATO DE AGRESSÃO É UMA GUERRA ILEGAL NÃO PROVOCADA E INJUSTIFICADA QUE PÕE EM CAUSA A INDEPENDÊNCIA SOBERANIA E INTEGRIDADE TERRITORIAL DA UCRÂNIA VIOLANDO FLAGRANTEMENTE O DIREITO INTERNACIONAL O AGREDIDO TEM O DIREITO DE SE DEFENDER E DEVE SER APOIADO NESSA LEGÍTIMA DEFESA A GUERRA DA RÚSSIA CONTRA A UCRÂNIA COLOCA EM QUESTÃO A ARQUITETURA DE SEGURANÇA EUROPEIA CONSTITUI A MAIS GRAVE SITUAÇÃO DE SEGURANÇA VIVIDA DESDE O FIM DA SEGUNDA GUERRA MUNDIAL E REPRESENTA UMA AMEAÇA AO CONJUNTO DOS PAÍSES DA EUROPA E DO ATLÂNTICO NORTE ESTES PAÍSES TÊM O DIREITO DE REFORÇAR A SUA PRÓPRIA CAPACIDADE DE DISSUASÃO E DEFESA E TÊM O DEVER MORAL E POLÍTICO DE AJUDAR A UCRÂNIA DEFENDENDO-SE A SI PRÓPRIA A UCRÂNIA DEFENDE-NOS A TODOS A TODOS OS QUE DEFENDEMOS OS VALORES DA LIBERDADE E DA DEMOCRACIA E QUEREMOS UMA ORDEM INTERNACIONAL BASEADA EM REGRAS E UMA PAZ ASSENTE NA CARTA DAS NAÇÕES UNIDAS EM QUE OS DIFERENDOS E OS CONFLITOS SÃO TRATADOS E RESOLVIDOS POR VIA DIPLOMÁTICA E	SENHOR PRESIDENTE DA REPÚBLICA PORTUGUESA SENHOR PRESIDENTE DA REPÚBLICA DA UCRÂNIA SENHOR PRIMEIRO-MINISTRO E DEMAIS MEMBROS DO GOVERNO SENHORAS E SENHORES DEPUTADOS SENHOR EMBAIXADOR DA UCRÂNIA EM PORTUGAL ILUSTRES CONVIDADOS MINHAS SENHORAS E MEUS SENHORES PORTUGAL CONDENOU DESDE O PRIMEIRO MOMENTO COM FIRME DETERMINAÇÃO A AGRESSÃO MILITAR DA FEDERAÇÃO RUSSA CONTRA A UCRÂNIA A MENOS DE TRÊS HORAS APÓS O SEU INÍCIO NA NOITE DE 23 PARA 24 DE FEVEREIRO O GOVERNO PORTUGUÊS REPROVOU-A PUBLICAMENTE NO DIA 24 TODOS OS ÓRGÃOS POLÍTICOS DE SOBERANIA O PRESIDENTE DA REPÚBLICA A ASSEMBLEIA DA REPÚBLICA E O GOVERNO CONDENAVAM EM UNÍSSONO O AGRESSOR E EXPRIMIAM SOLIDARIEDADE E APOIO AO AGREDIDO FIZERAM-A ENTÃO E TÊM REITERADAMENTE FEITO SEM QUALQUER HESITAÇÃO NEM AMBIGUIDADE PARA PORTUGAL O AGRESSOR É A FEDERAÇÃO RUSSA E O AGREDIDO É A UCRÂNIA O ATO DE AGRESSÃO É UMA GUERRA ILEGAL NÃO PROVOCADA E INJUSTIFICADA QUE PÕE EM CAUSA A INDEPENDÊNCIA SOBERANIA E INTEGRIDADE TERRITORIAL DA UCRÂNIA VIOLANDO FLAGRANTEMENTE O DIREITO INTERNACIONAL O AGREDIDO TEM O DIREITO DE SE DEFENDER E DEVE SER APOIADO NESSA LEGÍTIMA DEFESA A GUERRA DA RÚSSIA CONTRA A UCRÂNIA COLOCA EM QUESTÃO A ARQUITETURA DE SEGURANÇA EUROPEIA CONSTITUI A MAIS GRAVE SITUAÇÃO DE SEGURANÇA VIVIDA DESDE O FIM DA SEGUNDA GUERRA MUNDIAL E REPRESENTA UMA AMEAÇA AO CONJUNTO DOS PAÍSES DA EUROPA E DO ATLÂNTICO NORTE ESTES PAÍSES TÊM O DIREITO DE REFORÇAR A SUA PRÓPRIA CAPACIDADE DE DISSUASÃO E DEFESA E TÊM O DEVER MORAL E POLÍTICO DE AJUDAR A UCRÂNIA DEFENDENDO-SE A SI PRÓPRIA A UCRÂNIA DEFENDE-NOS A TODOS A TODOS OS QUE DEFENDEMOS OS VALORES DA LIBERDADE E DA DEMOCRACIA E QUEREMOS UMA ORDEM INTERNACIONAL BASEADA EM REGRAS E UMA PAZ ASSENTE NA CARTA DAS NAÇÕES UNIDAS EM QUE OS DIFERENDOS E OS CONFLITOS SÃO TRATADOS E RESOLVIDOS POR VIA DIPLOMÁTICA E

JUDICIAL E NÃO ATRAVÉS DA CHANTAGEM E DA AGRESSÃO PORTUGAL NÃO SE LIMITOU À CONDENAÇÃO DO AGRESSOR E À SOLIDARIEDADE COM O AGREDIDO FEZ CORRESPONDER OS ATOS ÀS PALAVRAS NO MESMÍSSIMO DIA 24 DE FEVEREIRO O NOSSO CONSELHO SUPERIOR DE DEFESA NACIONAL APROVOU SOB PROPOSTA DO GOVERNO E CONCORDÂNCIA DO COMANDANTE SUPREMO DAS FORÇAS ARMADAS AS MEDIDAS INDISPENSÁVEIS PARA REFORÇAR A NOSSA PARTICIPAÇÃO MILITAR NA DEFESA EUROPEIA E ATLÂNTICA E OS NOSSOS EMBAIXADORES DA UNIÃO EUROPEIA E DA NATO TRANSMITIRAM IMEDIATAMENTE A POSIÇÃO NACIONAL DE EMPENHAMENTO NAS MEDIDAS DE SANCIONAMENTO DA RÚSSIA E PROTEÇÃO DA UCRÂNIA AO MESMO TEMPO O PRIMEIRO-MINISTRO DECLARAVA QUE PORTUGAL ACOLHERIA TODOS OS CIDADÃOS UCRANIANOS EM NECESSIDADE DE PROTEÇÃO HUMANITÁRIA SEM QUALQUER RESTRIÇÃO SUBSEQUENTEMENTE O CONSELHO DE MINISTROS IMPLEMENTARIA UM MECANISMO EXCECIONAL DE REGULARIZAÇÃO IMEDIATA DA SITUAÇÃO DE QUALQUER PESSOA ORIUNDA DA UCRÂNIA DE MODO A GARANTIR-LHE O ACESSO PRONTO À PROTEÇÃO CIVIL SOCIAL E SANITÁRIA E A FACILITAR-LHE O EMPREGO E A INTEGRAÇÃO APOIÁMOS IMEDIATAMENTE A CONDENAÇÃO EXPRESSA PELAS NAÇÕES UNIDAS À AGRESSÃO RUSSA ESTIVEMOS NO PRIMEIRO GRUPO DE PAÍSES A SOLICITAR AO TRIBUNAL PENAL INTERNACIONAL A INVESTIGAÇÃO SOBRE OS CRIMES DE GUERRA COMETIDOS COOPERÁMOS NO ISOLAMENTO INTERNACIONAL DO REGIME DE PUTIN E ADVOGÁMOS E CONTINUAMOS A ADVOGAR SANÇÕES DURAS CONTRA OS RESPONSÁVEIS PELA AGRESSÃO E OS SETORES ECONÓMICOS INCLUINDO BANCA E ENERGIA QUE FINANCIAM A AGRESSÃO ENVIÁMOS E CONTINUAREMOS A ENVIAR BILATERALMENTE APOIO MILITAR HUMANITÁRIO E MATERIAL À UCRÂNIA E PARTICIPAMOS ATIVAMENTE NO ESFORÇO DA UNIÃO EUROPEIA MOBILIZANDO O MECANISMO EUROPEU DE APOIO À PAZ PARA PROVIDENCIAR À UCRÂNIA OS MEIOS DE DEFESA REFORÇÁMOS A NOSSA PARTICIPAÇÃO NO ROBUSTECIMENTO DA DEFESA EUROPEIA DESIGNADAMENTE NO QUADRO DA ALIANÇA ATLÂNTICA NA RESPOSTA PORTUGUESA À AGRESSÃO RUSSA CONTRA A UCRÂNIA PESOU CERTAMENTE O RELACIONAMENTO ESTREITO QUE EXISTE ENTRE OS DOIS PAÍSES O LAÇO MAIS FORTE É CONSTITUÍDO PELAS PESSOAS PELA COMUNIDADE UCRANIANA ESTABELECIDA EM PORTUGAL NA ORDEM DAS DEZENAS DE MILHARES DE PESSOAS BEM INTEGRADAS QUE EM MUITO CONTRIBUEM PARA A NOSSA ECONOMIA E EM CUJOS FILHOS SE ENCONTRAM ALGUNS DOS MELHORES ALUNOS DAS ESCOLAS PORTUGUESAS E PELAS FAMÍLIAS LUSO-UCRANIANAS QUE ENTRETANTO SE FORAM FORMANDO E RESIDEM QUER NUM QUER NO OUTRO PAÍS MAS SE O BOM RELACIONAMENTO BILATERAL POVO A POVO E ESTADO A ESTADO	JUDICIAL E NÃO ATRAVÉS DA CHANTAGEM E DA AGRESSÃO A UCRÂNIA NÃO SE LIMITOU À CONDENAÇÃO DO AGRESSOR E À SOLIDARIEDADE COM O AGREDIDO FEZ CORRESPONDER OS ATOS ÀS PALAVRAS NO MESMÍSSIMO DIA 24 DE FEVEREIRO O NOSSO CONSELHO SUPERIOR DE DEFESA NACIONAL APROVOU SUA PROPOSTA DE GOVERNO E CONCORDÂNCIA DO COMANDANTE SUPREMO DAS FORÇAS ARMADAS AS MEDIDAS INDISPENSÁVEIS PARA REFORÇAR A NOSSA PARTICIPAÇÃO MILITAR NA DEFESA EUROPEIA E ATLÂNTICA E OS NOSSOS EMBAIXADORES DA UNIÃO EUROPEIA E DA NATO TRANSMITIRAM IMEDIATAMENTE A POSIÇÃO NACIONAL DE EMPENHAMENTO NAS MEDIDAS DE SANCIONAMENTO DA RÚSSIA E PROTEÇÃO DA UCRÂNIA AO MESMO TEMPO O PRIMEIRO-MINISTRO DECLARAVA QUE PORTUGAL ACOLHERIA TODOS OS CIDADÃOS UCRANIANOS EM NECESSIDADE DE PROTEÇÃO HUMANITÁRIA SEM QUALQUER RESTRIÇÃO SUBSEQUENTEMENTE O CONSELHO-MINISTRO IMPLEMENTARIA UM MECANISMO EXCEPCIONAL DE REGULARIZAÇÃO IMEDIATA DA SITUAÇÃO DE QUALQUER PESSOA ORIUNDA DA UCRÂNIA DE MODO A GARANTIR-LHE O ACESSO PRONTO À PROTEÇÃO CIVIL SOCIAL E SANITÁRIA E A FACILITAR-LHE O EMPREGO E A INTEGRAÇÃO APOIAMOS IMEDIATAMENTE A CONDENAÇÃO EXPRESSA PELAS NAÇÕES UNIDAS À AGRESSÃO RUSSA ESTIVEMOS NO PRIMEIRO GRUPO DE PAÍSES A SOLICITAR AO TRIBUNAL PENAL INTERNACIONAL A INVESTIGAÇÃO SOBRE OS CRIMES DE GUERRA COMETIDOS COOPERAMOS NO ISOLAMENTO INTERNACIONAL DO REGIME DE PUTIN E ADVOGAMOS E CONTINUAMOS A ADVOGAR SANÇÕES DURAS CONTRA OS RESPONSÁVEIS PELA AGRESSÃO E OS SETORES ECONÓMICOS INCLUINDO BANCA E ENERGIA QUE FINANCIAM A AGRESSÃO ENVIAMOS E CONTINUAREMOS A ENVIAR BILATERALMENTE APOIO MILITAR HUMANITÁRIO E MATERIAL À UCRÂNIA E PARTICIPAMOS ATIVAMENTE NO ESFORÇO DA UNIÃO EUROPEIA MOBILIZANDO O MECANISMO EUROPEU DE APOIO À PAZ PARA PROVIDENCIAR À UCRÂNIA OS MEIOS DE DEFESA REFORÇAMOS A NOSSA PARTICIPAÇÃO NO ROBUSTECIMENTO DA DEFESA EUROPEIA DESIGNADAMENTE NO QUADRO DA ALIANÇA ATLÂNTICA NA RESPOSTA PORTUGUESA À AGRESSÃO RUSSA CONTRA A UCRÂNIA PESOU CERTAMENTE O RELACIONAMENTO ESTREITO QUE EXISTE ENTRE OS DOIS PAÍSES O LAÇO MAIS FORTE É CONSTITUÍDO PELAS PESSOAS PELA COMUNIDADE UCRANIANA ESTABELECIDA EM PORTUGAL NA ORDEM DAS DEZENAS DE MILHARES DE PESSOAS BEM INTEGRADAS QUE EM MUITO CONTRIBUEM PARA A NOSSA ECONOMIA E EM CUJOS FILHOS SE ENCONTRAM ALGUNS DOS MELHORES ALUNOS DAS ESCOLAS PORTUGUESAS E PELAS FAMÍLIAS LUSO-UCRANIANAS QUE ENTRETANTO SE FORAM FORMANDO E RESIDEM QUER NUM QUER NO OUTRO PAÍS MAS SE O BOM RELACIONAMENTO BILATERAL POVO A POVO E ESTADO A ESTADO

EXPLICA PARCIALMENTE A PRONTIDÃO E A CLAREZA DA REAÇÃO PORTUGUESA À AGRESSÃO DE QUE A UCRÂNIA É VÍTIMA ELA NÃO EXPLICA TUDO NEM O MAIS IMPORTANTE PORTUGAL É UM PAÍS MÉDIO À ESCALA EUROPEIA PEQUENO À ESCALA MUNDIAL NÃO É UMA POTÊNCIA DEMOGRÁFICA ECONÓMICA OU MILITAR MAS É MAS É UMA NAÇÃO COM HISTÓRIA COM UM POSICIONAMENTO GEOPOLÍTICO HÁ MUITO CONSOLIDADO E COM UMA POLÍTICA EXTERNA QUE NÃO VARIA COM O GOVERNO NO MOMENTO PORQUE EXPRIME INTERESSES NACIONAIS DURADOUROS ORA A CHAVE DA NOSSA POLÍTICA EXTERNA É O RESPEITO PELO DIREITO INTERNACIONAL A VINCULAÇÃO À CARTA DAS NAÇÕES UNIDAS A VALORIZAÇÃO DA PAZ E DA SEGURANÇA E O AMOR À LIBERDADE E É POR ISSO QUE ESTAMOS SEM HESITAÇÕES NEM AMBIGUIDADES PELA UCRÂNIA EM CUJO TERRITÓRIO SE TRAVA HOJE A LUTA PELA LIBERDADE A INDEPENDÊNCIA E A PAZ NA EUROPA SR PRESIDENTE VLADIMIR ZELENSKY É UMA HONRA PARA O PARLAMENTO PORTUGUÊS RECEBÊ-LO SOLENEMENTE E OUVIR AS SUAS PALAVRAS A PARTICIPAÇÃO DO PRESIDENTE E DO PRIMEIRO-MINISTRO DE PORTUGAL NESTA SESSÃO SOLENE MOSTRA BEM A UNIDADE NACIONAL EM TORNO DO APOIO À UCRÂNIA UM APOIO QUE JUNTA OS ÓRGÃOS DE SOBERANIA E QUE É PARTILHADO POR PARTIDOS POLÍTICOS DO GOVERNO E DA OPOSIÇÃO INDIGNADOS COM AS ATROCIDADES QUE ESTÃO A SER COMETIDAS E QUE V EXª ACABOU DE RELATAR EXEMPLIFICANDO CHORAMOS OS MORTOS CIVIS E MILITARES QUE TÊM SUCUMBIDO À BARBÁRIE E AO HORROR DA GUERRA INICIADA PELO REGIME DE PUTIN DEPLORAMOS A DESTRUIÇÃO SISTEMÁTICA E INTENCIONAL DE CIDADES INFRAESTRUTURAS HABITAÇÕES SAUDAMOS E ADMIRAMOS O ESFORÇO HEROICO DO EXÉRCITO E DA SOCIEDADE UCRANIANA NA DEFESA DA SUA PÁTRIA INCONGLUINDO NO DONBASS E APRESENTAMOS AS MAIS SENTIDAS CONDOLÊNCIAS POR TANTAS VIDAS INOCENTES JÁ PERDIDAS SABE V EXª SR PRESIDENTE DA UCRÂNIA QUE ENQUANTO ESTADO-MEMBRO DA UNIÃO EUROPEIA E DA NATO PORTUGAL SE BATE SEMPRE PELA PRESERVAÇÃO DA UNIDADE ESSENCIAL PARA A EFICÁCIA DAS NOSSAS DECISÕES E QUE NUNCA OBSTACULIZA ANTES FAVORECE OS PROCESSOS DE DECISÃO EM CURSO QUE VÃO NO SENTIDO DE APOIAR CADA VEZ MAIS O SEU PAÍS PREZAMOS AS ASPIRAÇÕES EUROPEIAS DA UCRÂNIA E TEMOS DEFENDIDO NÃO SÓ O REFORÇO DA COOPERAÇÃO NO QUADRO DO ACORDO DE ASSOCIAÇÃO JÁ EXISTENTE COMO TAMBÉM O EXAME PRONTO E ATENTO POR PARTE DAS INSTITUIÇÕES EUROPEIAS DO PEDIDO DE CANDIDATURA APRESENTADO PELA UCRÂNIA PERMITA-ME ENTRETANTO PRESIDENTE ZELENSKY QUE INDIVIDUALIZE A DIMENSÃO DA SOLIDARIEDADE E APOIO HUMANITÁRIO A QUE CALA MAIS FUNDO NA TRADIÇÃO HUMANISTA DO POVO PORTUGUÊS NO MOMENTO EM QUE FALO JÁ MAIS DE 31 MIL UCRANIANOS EM BUSCA DE	EXPLICA PARCIALMENTE A PRONTIDÃO E A CLAREZA DA REAÇÃO PORTUGUESA À AGRESSÃO DE QUE A UCRÂNIA É VÍTIMA PORTUGAL NÃO EXPLICA TUDO NEM O MAIS IMPORTANTE PORTUGAL É UM PAÍS MÉDIO À ESCALA EUROPEIA PEQUENO À ESCALA MUNDIAL NÃO É UMA POTÊNCIA DEMOGRÁFICA ECONÓMICA OU MILITAR MAS É UMA NAÇÃO COM HISTÓRIA COM UM POSICIONAMENTO GEOPOLÍTICO HÁ MUITO CONSOLIDADO E COM UMA POLÍTICA EXTERNA QUE NÃO VARIA COM O GOVERNO NO MOMENTO PORQUE EXPRIME INTERESSES NACIONAIS DURADOUROS ORA A CHAVE DA NOSSA POLÍTICA EXTERNA É O RESPEITO PELO DIREITO INTERNACIONAL A VINCULAÇÃO À CARTA DAS NAÇÕES UNIDAS A VALORIZAÇÃO DA PAZ E DA SEGURANÇA E O AMOR À LIBERDADE E É POR ISSO QUE ESTAMOS SEM HESITAÇÕES NEM AMBIGUIDADES PELA UCRÂNIA EM CUJO TERRITÓRIO SE TRAVA HOJE A LUTA PELA LIBERDADE A INDEPENDÊNCIA E A PAZ NA EUROPA SR PRESIDENTE VLADIMIR ZELENSKY É UMA HONRA PARA O PARLAMENTO PORTUGUÊS RECEBÊ-LO SOLENAMENTE E OUVIR AS SUAS PALAVRAS A PARTICIPAÇÃO DO PRESIDENTE E DO PRIMEIRO-MINISTRO DE PORTUGAL NESTA SESSÃO SOLENE MOSTRA BEM A UNIDADE NACIONAL EM TORNO DO APOIO À UCRÂNIA UM APOIO QUE JUNTA OS ÓRGÃOS DE SOBERANIA E QUE É PARTILHADO POR PARTIDOS POLÍTICOS DO GOVERNO E DA OPOSIÇÃO INDIGNADOS COM AS ATROCIDADES QUE ESTÃO A SER COMETIDAS E QUE V EXª ACABOU DE RELATAR EXEMPLIFICANDO CHORAMOS OS MORTOS CIVIS E MILITARES QUE TÊM SUCUMBIDO À BARBÁRIE E AO HORROR DA GUERRA INICIADA PELO REGIME DE PUTIN DEPLORAMOS A DESTRUIÇÃO SISTEMÁTICA E INTENCIONAL DE CIDADES INFRAESTRUTURAS HABITAÇÕES SAUDAMOS E ADMIRAMOS O ESFORÇO HEROICO DO EXÉRCITO E DA SOCIEDADE UCRANIANA NA DEFESA DA SUA PÁTRIA INCLUINDO NO DOMBÁSS E APRESENTAMOS AS MAIS SENTIDAS CONDOLÊNCIAS POR TANTAS VIDAS INOCENTES JÁ PERDIDAS SABE V EXª SR PRESIDENTE DA UCRÂNIA QUE ENQUANTO ESTANDO MEMBRO DA UNIÃO EUROPEIA DA NATO PORTUGAL SE BATE SEMPRE PELA PRESERVAÇÃO DA UNIDADE ESSENCIAL PARA A EFICÁCIA DAS NOSSAS DECISÕES E QUE NUNCA OBSTACULIZA ANTES FAVORECE OS PROCESSOS DE DECISÃO EM CURSO QUE VÃO NO SENTIDO DE APOIAR CADA VEZ MAIS O SEU PAÍS PREZAMOS AS ASPIRAÇÕES EUROPEIAS DA UCRÂNIA E TEMOS DEFENDIDO NÃO SÓ O REFORÇO DA COOPERAÇÃO NO QUADRO DO ACORDO DE ASSOCIAÇÃO JÁ EXISTENTE COMO TAMBÉM O EXAME PRONTO E ATENTO POR PARTE DAS INSTITUIÇÕES EUROPEIAS DO PEDIDO DE CANDIDATURA APRESENTADO PELA UCRÂNIA PERMITA-ME ENTRETANTO PRESIDENTE ZELENSKY QUE INDIVIDUALIZE A DIMENSÃO DA SOLIDARIEDADE E APOIO HUMANITÁRIO A QUE CALA MAIS FUNDO NA TRADIÇÃO HUMANISTA DO POVO PORTUGUÊS NO MOMENTO EM QUE FALO JÁ MAIS DE 31 MIL UCRANIANOS EM BUSCA DE

PROTEÇÃO HUMANITÁRIA FORAM ACOLHIDOS EM PORTUGAL E 2500 DAS VOSSAS CRIANÇAS FREQUENTAM AS NOSSAS ESCOLAS ESTE ACOLHIMENTO MOBILIZA TODOS OS PORTUGUESES GOVERNO E ADMINISTRAÇÃO CENTRAL REGIÕES AUTÓNOMAS E MUNICÍPIOS ORGANIZAÇÕES NÃO GOVERNAMENTAIS AS VÁRIAS CONFISSÕES RELIGIOSAS AS ESCOLAS AS EMPRESAS OS SINDICATOS E SOBRETUDO AS PESSOAS COMUNS AS PORTUGUESAS E OS PORTUGUESES ESTÃO EMPENHADOS NESTA VASTA CADEIA DE SOLIDARIEDADE E O TRATAMENTO QUE DEDICAM AOS UCRANIANOS EM NECESSIDADE E AUXÍLIO É AQUELE CARACTERÍSTICO DA NOSSA MANEIRA DE SER TRATAM-NOS COMO IGUAIS COMO IRMÃOS DA MESMA HUMANIDADE UM POUCO POR TODO O PAÍS MILHARES E MILHARES DE VOLUNTÁRIOS TÊM VINDO A PROVIDENCIAR TRANSPORTE ALOJAMENTO EMPREGO E INTEGRAÇÃO EM ESTREITA COLABORAÇÃO COM A EMBAIXADA DA UCRÂNIA EM PORTUGAL COM AS ASSOCIAÇÕES REPRESENTATIVAS DA COMUNIDADE UCRANIANA E COM OS SEUS COMPATRIOTAS JÁ AQUI ESTABELECIDOS A SENHORA EMBAIXADORA E VÁRIOS REPRESENTANTES DA COMUNIDADE DÃO-NOS ALIÁS O GOSTO DE ASSISTIR A ESTA SAÚDA A ESTA SESSÃO E A TODOS DESEJO-VOS SAUDAR SENHOR PRESIDENTE DA REPÚBLICA DA UCRÂNIA OUVIMOS COM TODA A ATENÇÃO E DE ESPÍRITO ABERTO AS SUAS PALAVRAS E EM PARTICULAR OS SEUS APELOS NO ORDENAMENTO CONSTITUCIONAL PORTUGUÊS É AO GOVERNO QUE COMPETE CONDUZIR A POLÍTICA EXTERNA E BASTA NOTAR O NÍVEL DE REPRESENTAÇÃO DO GOVERNO NESTA SESSÃO LIDERADA PELO PRIMEIRO-MINISTRO PARA SE COMPREENDER QUE AS PROPOSTAS E PEDIDOS DE V EXª SR PRESIDENTE DA UCRÂNIA SERÃO BEM EXAMINADOS NA SUA FUNÇÃO DE FISCALIZAÇÃO AS SENHORAS E OS SENHORES DEPUTADOS ACOMPANHARÃO TAMBÉM DE PERTO AS DECISÕES DO GOVERNO MAS POSSO DESDE JÁ ASSEGURAR-LHE PRESIDENTE ZELENSKY QUE CONTA COM PORTUGAL CONTA COM A NOSSA DEFESA INTRANSIGENTE DAS LEIS QUE REGULAM AS RELAÇÕES INTERNACIONAIS E DO DIREITO À INDEPENDÊNCIA E SOBERANIA NACIONAL CONTA COM O NOSSO EMPENHAMENTO DESIGNADAMENTE NO QUADRO DA UNIÃO EUROPEIA E DA NATO NA DEFESA DA LIBERDADE EM TODOS OS TERRITÓRIOS DA EUROPA NO SANCIONAMENTO CADA VEZ MAIS INTENSO DO AGRESSOR E NO APOIO NECESSÁRIO AO AGREDIDO NA GUERRA DA RÚSSIA CONTRA A UCRÂNIA CONTA COM A SOLIDARIEDADE E A AÇÃO EFETIVA DO POVO E DAS AUTORIDADES PORTUGUESAS NOMEADAMENTE NO CAMPO HUMANITÁRIO E NO ACOLHIMENTO E INTEGRAÇÃO DAS FAMÍLIAS DE MIGRANTES E REFUGIADOS E CONTA COM TODO O NOSSO APOIO AOS SEUS ESFORÇOS SR PRESIDENTE ZELENSKY PARA ENCONTRAR OS CAMINHOS DE UMA PAZ BASEADA NA RECUSA DA AGRESSÃO E NA SOLUÇÃO POLÍTICA NEGOCIADA PARA OS DIFERENDOS COMO V EXª PRESIDENTE ZELENSKY BEM SABE	PROTEÇÃO HUMANITÁRIA FORAM ACOLHIDOS EM PORTUGAL E 2500 DAS VOSSAS CRIANÇAS FREQUENTAM AS NOSSAS ESCOLAS ESTE ACOLHIMENTO MOBILIZA TODOS OS PORTUGUESES GOVERNO E ADMINISTRAÇÃO CENTRAL REGIÕES AUTÓNOMAS E MUNICÍPIOS ORGANIZAÇÕES NÃO GOVERNAMENTAIS AS VÁRIAS CONFISSÕES RELIGIOSAS AS ESCOLAS AS EMPRESAS OS SINDICATOS E SOBRETUDO AS PESSOAS COMUNS AS PORTUGUESAS E OS PORTUGUESES ESTÃO EMPENHADOS NESTA VASTA CADEIA DE SOLIDARIEDADE E O TRATAMENTO QUE DEDICAM AOS UCRANIANOS EM NECESSIDADE E AUXÍLIO É AQUELE CARACTERÍSTICO DA NOSSA MANEIRA DE SER TRATAM-NOS COMO IGUAIS COMO IRMÃOS DA MESMA HUMANIDADE UM POUCO POR TODO O PAÍS MILHARES E MILHARES DE VOLUNTÁRIOS TÊM VINDO A PROVIDENCIAR TRANSPORTE ALOJAMENTO EMPREGO E INTEGRAÇÃO EM ESTREITA COLABORAÇÃO COM A EMBAIXADA DA UCRÂNIA EM PORTUGAL COM AS ASSOCIAÇÕES REPRESENTATIVAS DA COMUNIDADE UCRANIANA E COM OS SEUS COMPATRIOTAS JÁ AQUI ESTABELECIDOS A SENHORA EMBAIXADORA E VÁRIOS REPRESENTANTES DA COMUNIDADE DÃO-NOS ALIÁS O GOSTO DE ASSISTIR A ESTA SESSÃO E A TODOS DESEJO-VOS SAUDAR SENHOR PRESIDENTE DA REPÚBLICA DA UCRÂNIA OUVIMOS COM TODA A ATENÇÃO E DE ESPÍRITO ABERTO AS SUAS PALAVRAS E EM PARTICULAR OS SEUS APELOS NO ORDENAMENTO CONSTITUCIONAL PORTUGUÊS É AO GOVERNO QUE COMPETE CONDUZIR A POLÍTICA EXTERNA E BASTA NOTAR O NÍVEL DE REPRESENTAÇÃO DO GOVERNO NESTA SESSÃO LIDERADA PELO PRIMEIRO MINISTRO PARA SE COMPREENDER QUE AS PROPOSTAS E PEDIDOS DE V EXª SR PRESIDENTE DA UCRÂNIA SERÃO BEM EXAMINADOS NA SUA FUNÇÃO DE FISCALIZAÇÃO AS SENHORAS E OS SENHORES DEPUTADOS ACOMPANHARÃO TAMBÉM DE PERTO AS DECISÕES DO GOVERNO MAS POSSO DESDE JÁ ASSEGURAR-LHE PRESIDENTE ZELENSKY QUE CONTA COM PORTUGAL CONTA COM A NOSSA DEFESA INTRANSIGENTE DAS LEIS QUE REGULAM AS RELAÇÕES INTERNACIONAIS E DO DIREITO À INDEPENDÊNCIA E SOBERANIA NACIONAL CONTA COM O NOSSO EMPENHAMENTO DESIGNADAMENTE NO QUADRO DA UNIÃO EUROPEIA E DA NATO NA DEFESA DA LIBERDADE EM TODOS OS TERRITÓRIOS DA EUROPA NO SANCIONAMENTO CADA VEZ MAIS INTENSO DO AGRESSOR E NO APOIO NECESSÁRIO AO AGREDIDO NA GUERRA DA RÚSSIA CONTRA A UCRÂNIA CONTA COM A SOLIDARIEDADE E A AÇÃO EFETIVA DO POVO E DAS AUTORIDADES PORTUGUESAS NOMEADAMENTE NO CAMPO HUMANITÁRIO E NO ACOLHIMENTO E INTEGRAÇÃO DAS FAMÍLIAS DE MIGRANTES E REFUGIADOS E CONTA COM TODO O NOSSO APOIO AOS SEUS ESFORÇOS SR PRESIDENTE ZELENSKY PARA ENCONTRAR OS CAMINHOS DE UMA PAZ BASEADA NA RECUSA DA AGRESSÃO E NA SOLUÇÃO POLÍTICA NEGOCIADA PARA OS DIFERENTES COMO V EX PRESIDENTE ZELENSKY BEM SABE

PORTUGAL É JUSTO TÍTULO CONSIDERADO COMO UM DOS PAÍSES MAIS PACÍFICOS DO MUNDO É O NOSSO MODO HUMANISTA DE CONCEBER AS RELAÇÕES ENTRE OS POVOS E AS NAÇÕES NÓS APRECIAMOS AS VIAGENS O COMÉRCIO A COMUNICAÇÃO COOPERAÇÃO E AS DESCOBERTAS QUE VAMOS FAZENDO DAS CULTURAS UNS DOS OUTROS TEMOS MUITO ORGULHO EM DISPORMOS DESDE 2019 NUMA PRAÇA DE LISBOA DO BUSTO DO VOSSO POETA NACIONAL TARAS TCHEPCHENKO E RECORDAMOS COM EMOÇÃO O ENCONTRO NOS ANOS DA GRANDE GUERRA NO NORTE DE PORTUGAL ENTRE SÓNIA DELAUNAY NASCIDA SARAH STERN EM GRADIZHSK NA UCRÂNIA E ENTÃO EM FUGA DA GUERRA E O NOSSO PINTOR AMADEUS SOUSA CARDOSO O ENCONTRO DE DUAS FIGURAS MAIORES DA REVOLUÇÃO MODERNISTA NA ARTE EUROPEIA MAS NÃO SOMOS INGÉNUOS PARA VOLTARMOS À PAZ QUE PERMITE E ESTIMULA O DESENVOLVIMENTO DOS LAÇOS CULTURAIS PRECISAMOS DE GANHAR A PAZ E PARA GANHAR A PAZ PRECISAMOS DE FAZER FRENTE À AGRESSÃO E DE FORÇAR O AGRESSOR A PARAR A AGRESSÃO ENVOLVENDO-SE NUM PROCESSO NEGOCIAL SÉRIO CONDUCENTE À PAZ NESSE PONTO ESTAMOS POR ISSO EM NOME DO PARLAMENTO PORTUGUÊS E NA PRESENÇA CONCORDANTE DO PRESIDENTE DA REPÚBLICA E DO PRIMEIRO-MINISTRO DE PORTUGAL ME PERMITO DIRIGIR-ME A V EXª SR PRESIDENTE DA UCRÂNIA PARA LHE DIZER QUE A LUTA DO SEU PAÍS PELA LIBERDADE É A LUTA DA EUROPA TODA PELA LIBERDADE E A ESSA LUTA PELA LIBERDADE QUE O PORTUGAL DEMOCRÁTICO NUNCA FALHOU NÃO FALTA E NÃO FALTARÁ MUITO OBRIGADO SR PRESIDENTE DA UCRÂNIA.	PORTUGAL É JUSTO TÍTULO CONSIDERADO COMO UM DOS PAÍSES MAIS PACÍFICOS DO MUNDO É O NOSSO MODO HUMANISTA DE CONCEBER AS RELAÇÕES ENTRE OS POVOS E AS NAÇÕES NÓS APRECIAMOS AS VIAGENS O COMÉRCIO A COMUNICAÇÃO A COOPERAÇÃO E AS DESCOBERTAS QUE VAMOS FAZENDO DAS CULTURAS UNS DOS OUTROS TEMOS MUITO ORGULHO EM DISPOR-NOS DESDE 2019 NUMA PRAÇA DE LISBOA DO BUSTO DO VOSSO POETA NACIONAL TARAS SHEVCHENKO E RECORDAMOS COM EMOÇÃO O ENCONTRO NOS ANOS DA GRANDE GUERRA NO NORTE DE PORTUGAL ENTRE SONIA DELONAY NASCIDA SARA STERN EM GRAZYSK NA UCRÂNIA E ENTÃO EM FUGA DA GUERRA E O NOSSO PINTOR AMADEO DE SOUSA-CARDOSO O ENCONTRO DE DUAS FIGURAS MAIORES DA REVOLUÇÃO MODERNISTA NA ARTE EUROPEIA MAS NÃO SOMOS INGÉNUOS PARA VOLTARMOS À PAZ QUE PERMITE E ESTIMULA O DESENVOLVIMENTO DOS LAÇOS CULTURAIS PRECISAMOS DE GANHAR A PAZ E PARA GANHAR A PAZ PRECISAMOS DE FAZER FRENTE À AGRESSÃO E DE FORÇAR O AGRESSOR A PARAR A AGRESSÃO ENVOLVENDO-SE NUM PROCESSO NEGOCIAL SÉRIO CONDUZENTE À PAZ NESSE PONTO ESTAMOS POR ISSO EM NOME DO PARLAMENTO PORTUGUÊS E NA PRESENÇA CONCORDANTE DO PRESIDENTE DA REPÚBLICA E DO PRIMEIRO-MINISTRO DE PORTUGAL ME PERMITO DIRIGIR-ME A V EX PRESIDENTE DA UCRÂNIA PARA LHE DIZER QUE A LUTA DO SEU PAÍS PELA LIBERDADE É A LUTA DA EUROPA TODA PELA LIBERDADE E É ESSA LUTA PELA LIBERDADE QUE O PORTUGAL DEMOCRÁTICO NUNCA FALHOU NÃO FALTA E NÃO FALTARÁ MUITO OBRIGADO SR PRESIDENTE DA UCRÂNIA.

Anexo III - Exemplos de erros comuns em transcrições com Whisper

A Tabela 13 identifica os erros mais comuns do Whisper na transcrição que se encontra no Anexo II - Análise comparativa de transcrições, e onde se verificam as seguintes questões:

- A vermelho, na 1.ª coluna, palavras e expressões mal transcritas pelo Whisper, com a respetiva correção, a verde, na 2.ª coluna;
- A laranja, na 1.ª coluna, palavras e expressões corretamente transcritas, mas que precisam de correção devido à má pronúncia/dicção do orador ou à incorreta pontuação, a verde, na 2.ª coluna;
- A verde, na 1ª coluna, expressões repetidas ou incorretas corrigidas pelo Whisper, na 2.ª coluna a laranja.

Tabela 13 - Exemplos de erros comuns em transcrições com Whisper

	Transcrição Whisper	Transcrição manual
1	A menos de três horas após o seu início,	Menos de três horas após o seu início,
2	Fizeram-a então, e têm reiteradamente feito,	Fizeram-no então, e têm reiteradamente feito,
3	A Ucrânia não se limitou à condenação do agressor	Portugal não se limitou à condenação do agressor
4	Conselho Superior de Defesa Nacional aprovou sua proposta	Conselho Superior de Defesa Nacional aprovou, sob
5	proposta de governo e concordância	proposta do Governo e concordância
6	Subsequentemente, o Conselho-Ministro	Subsequentemente, o Conselho de Ministros
7	implementaria um mecanismo excepcional	implementaria um mecanismo excecional
8	Apoiamos imediatamente a condenação expressa	Apoiámos imediatamente a condenação expressa
9	Advogamos e continuamos a advogar	Advogámos e continuamos a advogar
10	enviamos e continuaremos a enviar	enviámos e continuaremos a enviar
11	participamos ativamente no Esforço da União Europeia	participamos ativamente no esforço da União Europeia
12	Reforçamos a nossa participação	Reforçámos a nossa participação
13	Portugal não explica tudo	Ela [ele] não explica tudo
14	Mas é uma nação com história	Mas é… Mas é uma nação com história
15	que não varia com o governo no momento	que não varia com o Governo no momento
16	Vladimir Zelensky, é uma honra	Volodymir Zelenskyy, é uma honra
17	o Parlamento Português	o Parlamento português
18	recebê-lo solenamente	recebê-lo solenemente
19	partidos políticos do governo e da oposição	partidos políticos do Governo e da oposição
20	Incluindo no Dómbass	Incongluindo… incluindo no Donbass
21	que enquanto estando membro da União Europeia	que enquanto Estado-Membro da União Europeia
22	, no momento em que falo	. No Momento em que falo
23	Governo e administração central	Governo e Administração central
24	o gosto de assistir a esta sessão	o gosto de assistir a esta sauda… a esta sessão
25	liderada pelo Primeiro Ministro	liderada pelo Primeiro-Ministro
26	negociada para os diferentes.	negociada para os diferendos.
27	Como V. Ex. Presidente	Como V. Ex.ª Presidente
28	Zelensky	Zelenskyy
29	orgulho em dispor-nos, desde 2019	orgulho em dispormos, desde 2019
30	Taras Tchepchenko	Taras Shevchenko
31	Sónia Delonay	Sonia Delaunay
32	nascida Sara Stern	nascida Sarah Stern
33	em Grazysk	em Gradizhsk
34	Amadeus Sousa Cardoso	Amadeo de Souza-Cardoso
35	conduzente à paz	conducente à paz
36	dirigir-me a V. Ex.ª	dirigir-me a V. Ex.ª
37	E é essa luta	E, a essa luta,

Anexo IV - Inquérito sobre o Sistema de Transcrição Automática (STAAR)

Este inquérito pretende recolher a sua opinião relativamente ao sistema de transcrições automáticas de debates parlamentares (STAAR). A sua opinião é essencial para avaliar a eficácia e desempenho do STAAR, bem como adaptá-lo ainda mais às necessidades do ambiente parlamentar.

Os dados recolhidos são anónimos, tratados com confidencialidade e servem apenas para fins estatísticos. O preenchimento do inquérito demora entre 3 e 5 minutos.

Caracterização

C1: Serviço
- Divisão de Redação | Divisão de Apoio às Comissões | Outro

C2: Idade (em anos)
- Menos de 29 | Entre 30 e 39 | Entre 40 e 49 | Entre 50 e 59 | Mais de 60

C3: Experiência nas funções atuais (em anos)
- Menos de 9 | Entre 10 a 19 | Entre 20 a 29 | Entre 30 e 39 | Mais de 40

Questões sobre experiência com o Sistema de Transcrição Automático (STAAR)

Q1: Com que frequência utiliza os documentos resultantes da transcrição automática?
- Diariamente | Semanalmente | Mensalmente | Raramente

Q2: Como classifica a facilidade de uso das transcrições geradas pelo STAAR?
- Muito fácil | Fácil | Neutro | Difícil | Muito difícil

Q3: Como avalia a rapidez com que as transcrições são disponibilizadas?
- Muito rápida | Rápida | Neutro | Lenta | Muito lenta

Q4: Como avalia a precisão das transcrições geradas pelo STAAR?
- Muito alta | Alta | Aceitável | Baixa | Muito baixa

Q5: Considera que a separação de texto por orador é útil?
- Sim, muito | Sim, um pouco | Neutro | Não muito | Não, de todo

Q6: A formatação do texto e identificação de correções automáticas facilita o seu trabalho?
- Sim, muito | Sim, um pouco | Neutro | Não muito | Não, de todo

Q7: Considera que o STAAR atende às suas necessidades relativas à transcrição dos áudios?
- Totalmente | Parcialmente | Neutro | Raramente | Nunca

Q8: Considera a implementação do STAAR como impacto positivo na função que desempenha?
- Sim, muito | Sim, um pouco | Neutro | Não muito | Não, de todo

Q9: Escolha duas funcionalidades adicionais que gostaria de ver implementadas no STAAR?
- Criar um interface para adicionar/remover palavras de substituição automática;
- Desenvolver um modelo linguístico próprio a partir de registos da AR para reduzir a taxa de erro nas transcrições;
- Disponibilizar num só interface o áudio, o texto transcrito e pedaleira/atalhos para controlo do áudio;
- Gerar com maior rapidez os ficheiros de transcrição;
- Identificar o nome/partido exato do orador no texto transcrito;
- Mudar o formato da transcrição para formato estruturado, para ser usado por diversas aplicações;
- Produzir resumos/sumários da transcrição de forma automática;
- Verificar os nomes de Deputados em funções e corrigir nomes muito semelhantes que aparecem na transcrição;
- Outra opção…

Anexo V - Resultados do inquérito ao STAAR

Este anexo detalha os resultados estatísticos obtidos a partir das 22 respostas ao inquérito realizado, de forma a apresentar uma análise quantitativa das avaliações dos inquiridos sobre o STAAR.

Caracterização

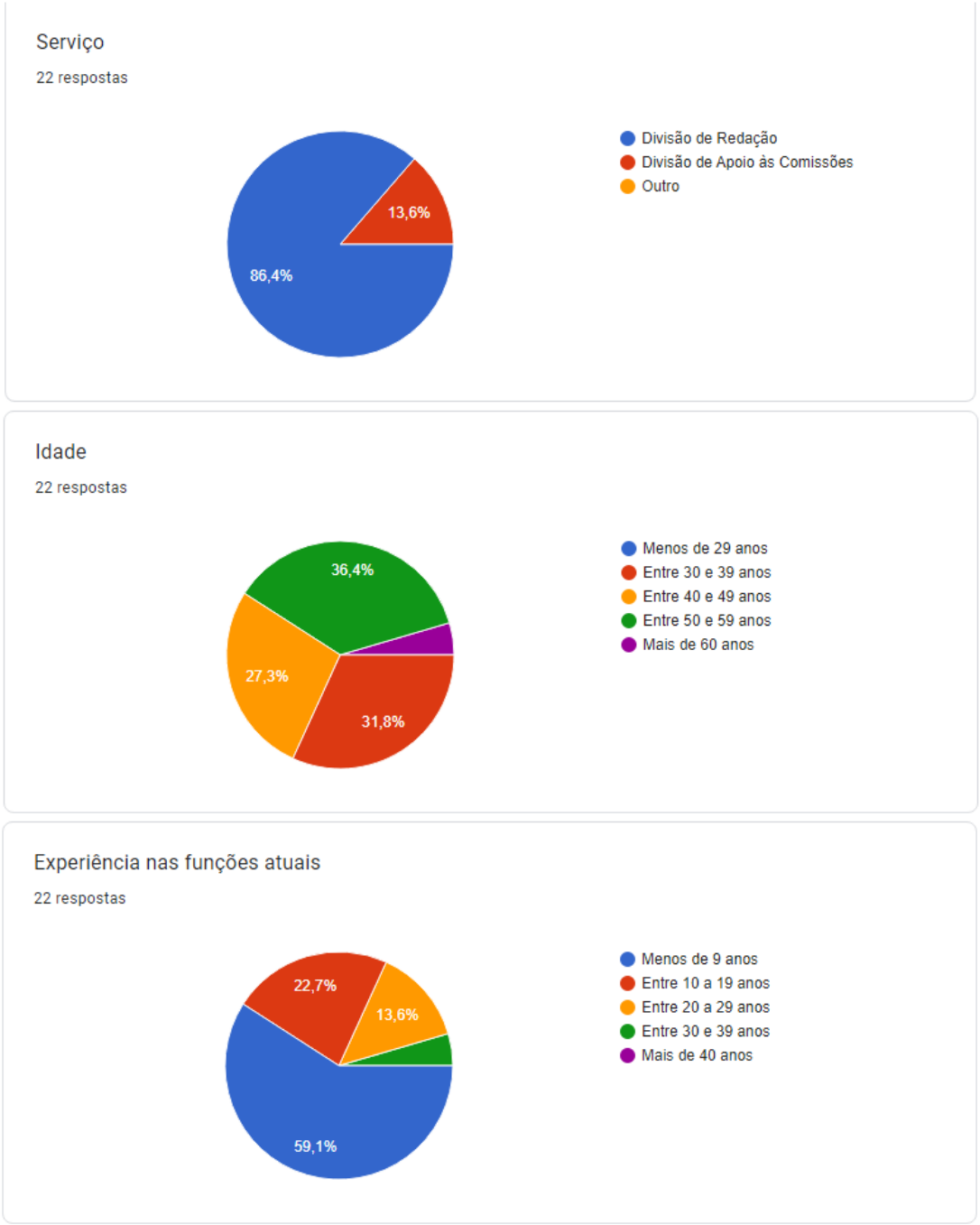


Figura 38 - Resposta ao Inquérito - Caracterização

Questões sobre experiência com o Sistema de Transcrição Automático (STAAR)

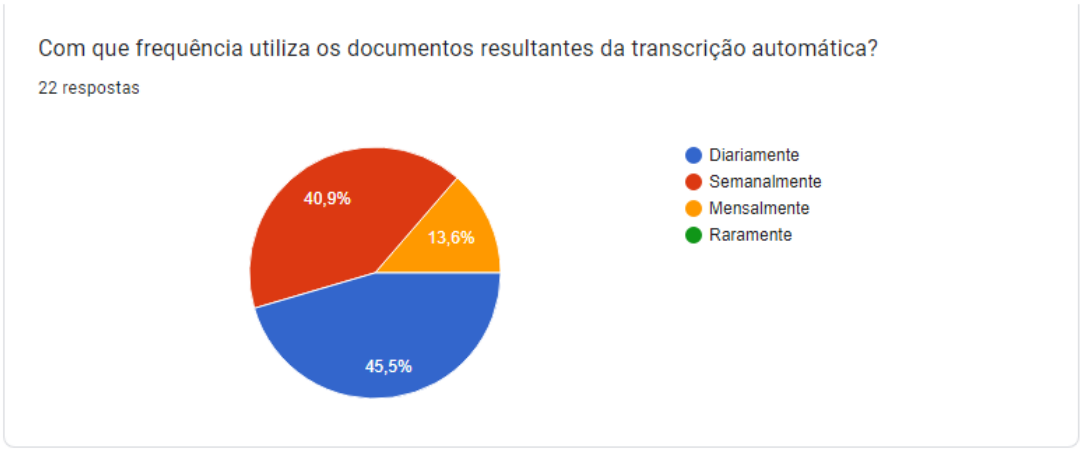


Figura 39 - Resposta ao Inquérito - Frequência de utilização

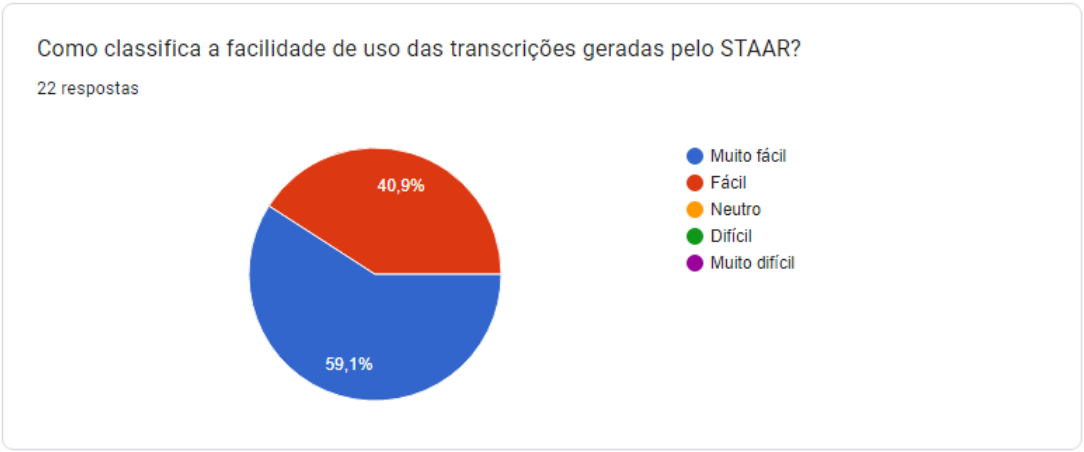


Figura 40 - Resposta ao Inquérito - Facilidade de uso

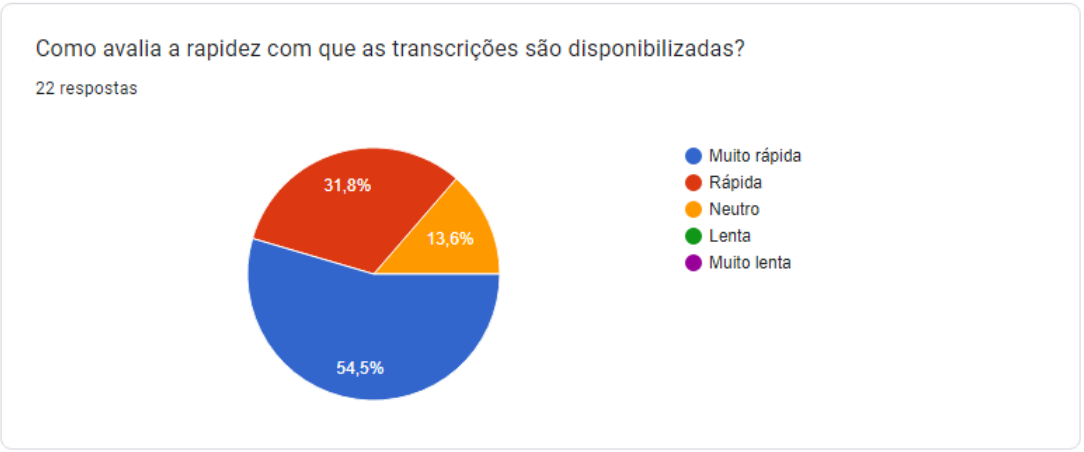


Figura 41 - Resposta ao Inquérito - Rapidez de transcrição

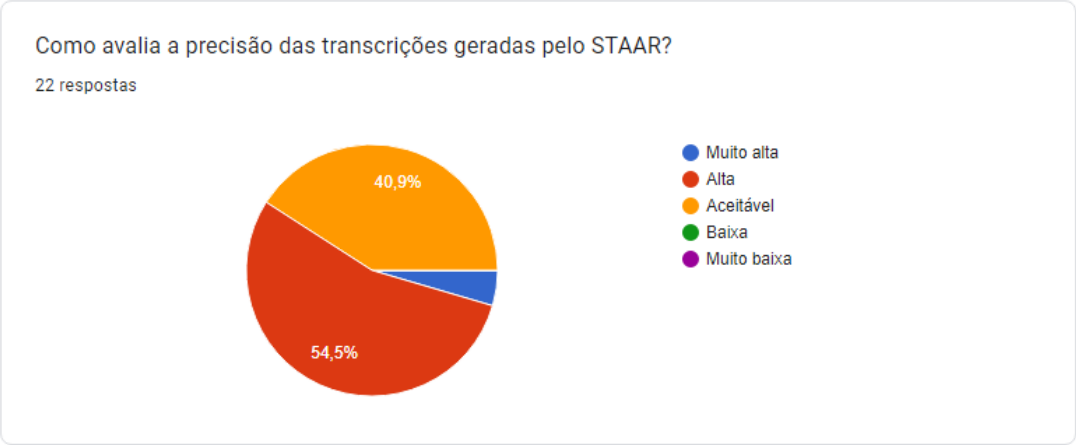


Figura 42 - Resposta ao Inquérito - Precisão das transcrições

Figura 43 - Resposta ao Inquérito - Separação de texto por orador

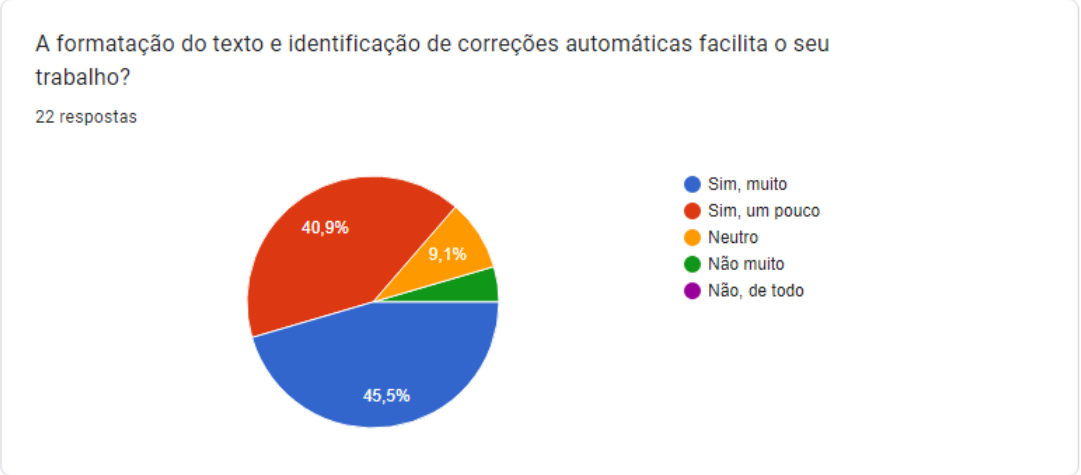


Figura 44 - Resposta ao Inquérito - Formatação do texto

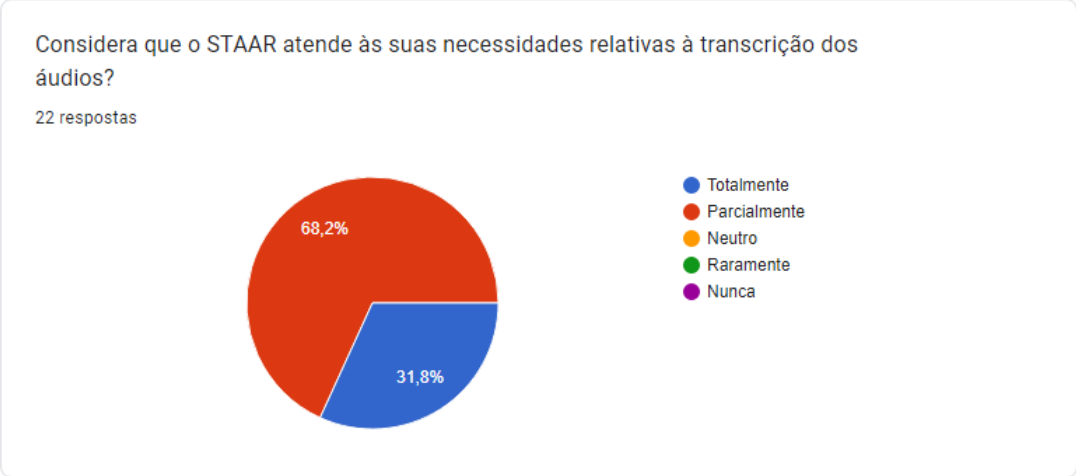


Figura 45 - Resposta ao Inquérito - Responde às necessidades

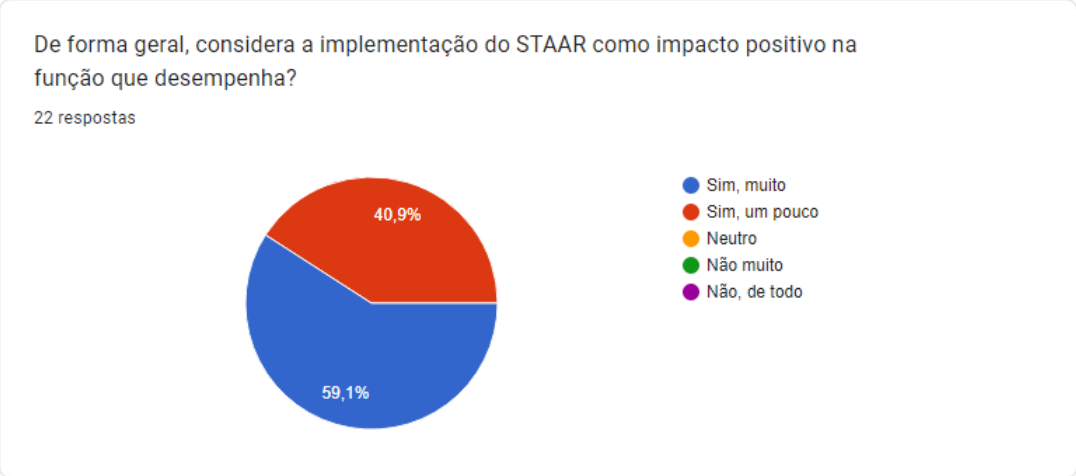


Figura 46 - Resposta ao Inquérito - Impacto positivo na função

Figura 47 - Resposta ao Inquérito - Funcionalidades a implementar