

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE ARTES E LETRAS
CURSO DE BACHARELADO EM LETRAS PORTUGUÊS/LITERATURAS

Sara Aparecida Gomes

**ANÁLISE DE ASPECTOS LINGÜÍSTICOS DA ASSISTENTE VIRTUAL
PARA RECONHECIMENTO DE VOZ IARA**

Santa Maria, RS
2023

Sara Aparecida Gomes

**ANÁLISE DE ASPECTOS LINGUÍSTICOS DA ASSISTENTE
VIRTUAL PARA RECONHECIMENTO DE VOZ IARA**

Trabalho de conclusão de curso,
apresentado ao Curso de Bacharelado em
Letras Português/Literaturas, da
Universidade Federal de Santa Maria
(UFSM, RS) – Campus Santa Maria,
como requisito parcial para obtenção do
título de **Bacharel em Letras**.

Orientador: Profa. Dra. Tatiana Keller

Co-orientador: Prof. Dr. Magnun Rochel Madruga

Santa Maria, RS
2023

Sara Aparecida Gomes

**ANÁLISE DE ASPECTOS LINGUÍSTICOS DA ASSISTENTE
VIRTUAL PARA RECONHECIMENTO DE VOZ IARA**

Trabalho de conclusão de curso,
apresentado ao Curso de Bacharelado em
Letras Português/Literaturas, da
Universidade Federal de Santa Maria
(UFSM, RS) – Campus Santa Maria,
como requisito parcial para obtenção do
título de **Bacharela em Letras**.

Aprovado em (dia) de fevereiro de 2023:

Membro 1 da banca de avaliação do trabalho (UFSM)
(Presidente/Orientador)

Membro 2 da banca de avaliação do trabalho (UFSM)

Membro 3 da banca de avaliação do trabalho (UFSM)

Santa Maria, RS
2023

RESUMO

ANÁLISE DE ASPECTOS LINGUÍSTICOS DA ASSISTENTE VIRTUAL PARA RECONHECIMENTO DE VOZ IARA

AUTOR: Sara Aparecida Gomes
ORIENTADORA: Tatiana Keller (UFSM)
CO-ORIENTADOR: Magnun Rochel Madruga (UFMG)

O Reconhecimento de Voz (RV) está presente em diversas Inteligências Artificiais (IA). As IAs auxiliam a vida dos usuários ao fazer algumas tarefas do cotidiano humano, como ligar aparelhos eletrônicos, enviar mensagens de textos, entre outras. Desde 2018, no Brasil, a *Iara Health*, uma *startup* brasileira, desenvolveu a IA Iara para auxiliar o trabalho de profissionais da saúde, médicos radiologistas. A partir de uma análise da estrutura linguística, especialmente da sílaba, este trabalho propõe-se a analisar os fatores estruturais que mais impactam o RV da *Iara Health*, comprometendo seu desempenho. Verificou-se que as cinco principais estruturas linguísticas não reconhecidas pela Iara são: *onset* complexo (com 11 ocorrências); similaridade fonológica (com 10 ocorrências); coda nasal (com 8 ocorrências); e flexão de número e epêntese (ambas com 7 ocorrências cada). Na comparação com a Google Assistente, os resultados indicam que a IA da empresa de tecnologia americana transcreveu de forma incorreta 143 palavras das 509 totais do *corpus* dessa pesquisa, o que corresponde a, aproximadamente, 28% de palavras erradas em todo o *corpus*. Já a IA da *startup* brasileira errou menos de 9% das palavras do *corpus*.

Palavras-chave: Fonética. Fonologia. Linguística. Inteligência Artificial. Reconhecimento de Voz. Iara Health.

Sumário

1 INTRODUÇÃO	6
2 FUNDAMENTAÇÃO TEÓRICA.....	7
2.1 O QUE É UMA INTELIGÊNCIA ARTIFICIAL?	7
2.2 O QUE É RECONHECIMENTO DE VOZ?	7
2.3 A <i>IARA HEALTH</i>	8
2.4 POR QUE OS RADIOLOGISTAS?	9
2.5 A SÍLABA NO PORTUGUÊS BRASILEIRO	10
2.5.1 Onset silábico.....	11
2.5.2 A Coda Silábica	12
3 METODOLOGIA.....	13
4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	13
4.1 RESULTADOS DA PERFORMANCE DA IARA	17
4.2. COMPARAÇÃO IARA VERSUS GOOGLE ASSISTENTE.....	20
5. CONSIDERAÇÕES FINAIS	22
REFERÊNCIAS	23
ANEXOS	25

1 INTRODUÇÃO

Cada vez mais é comum ouvir falar em Reconhecimento de Voz e Inteligência Artificial. Isso porque, progressivamente, essas tecnologias foram adentrando o cotidiano da humanidade. O que, até um tempo atrás, era algo real apenas em filmes de ficção científica, atualmente está disponível e faz parte do dia a dia da população, seja na palma da mão de qualquer usuário que tenha um *smartphone*, um *notebook* ou nas casas que possuam um *echo dot*, uma *smart TV*, entre outros dispositivos eletrônicos.

Essa realidade é justificável em virtude do desenvolvimento diverso e veloz que os ramos da ciência – tanto nas áreas de Matemática, Computação, Engenharia quanto nas de Linguística, Psicologia – vêm alcançando, recentemente. Esse avanço atingiu diversas áreas, seja na educação, atendimento ao público e até na área da saúde, sobre a qual se detém este trabalho. Especificamente, analisamos, neste estudo, a Assistente Virtual da *Iara Health*.

A *Iara Health* é uma *startup* brasileira que oferece aos médicos uma Inteligência Artificial capaz de transcrever as análises feitas de cada exame realizado, por meio de Reconhecimento de Voz (doravante RV) que mapeia a fala em texto.

Para a análise da performance da *Iara*, foram selecionados 33 textos, os quais apresentaram pelo menos uma palavra transcrita de forma incorreta pela rede em questão. Em virtude disso, objetiva-se compreender quais os fatores linguísticos que possam estar comprometendo a eficiência da *Iara* ao realizar a sua função.

Esta pesquisa enquadra-se na área de Letras e fará uma análise fonética e fonológica de textos selecionados da área médica. Destaca-se que para esse estudo não será feita uma distinção entre Fonética e Fonologia por tratar-se de um trabalho relativo à transposição de fala em escrita.

Para isso, esse estudo está dividido em: Introdução, que contextualiza a pesquisa; Fundamentação Teórica, a qual apresenta os conceitos essenciais para a compreensão do trabalho e de seus resultados, como a definição de Inteligência Artificial, de Reconhecimento de Voz e de sílaba; Metodologia, em que é descrita como foi feita a elaboração do *corpus* e o desenvolvimento da pesquisa; Apresentação e análise dos resultados, em que será mostrado e analisado cada um dos casos obtidos na pesquisa; e Considerações finais, na qual será exposto o que foi possível descobrir, em

virtude da pesquisa, ou seja, quais foram os achados proporcionadas por meio desse estudo.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 O QUE É UMA INTELIGÊNCIA ARTIFICIAL?

De acordo com Gomes (2010), a Inteligência Artificial (IA) é um ramo da Ciência da Computação que tem como principal interesse fazer com que os computadores “pensem” ou “ajam” de forma inteligente. Ainda segundo o autor, a história da IA começou entre as décadas de 40 e 50 e continua atualmente.

Segundo Gomes, a Inteligência Artificial não está restrita à Ciência da Computação, pelo contrário, ela também está relacionada com diversos campos da ciência, entre eles pode-se ressaltar a Biologia, a Linguística, a Psicologia.

A principal justificativa para o crescimento das Inteligências Artificiais é o desenvolvimento de novas tecnologias para a extração, o processamento, o armazenamento e a transmissão de dados (CARVALHO, 2021). Segundo a professora Leilane Nunes de Barros (2016), a IA pode ser considerada como um estudo/uma construção de sistemas computacionais inteligentes que “pensam” e que “agem” de forma racional.

Gomes cita algumas aplicações elaboradas por esse sistema, dentre eles destaca-se o Processamento de Linguagem Natural (PLN) que possibilita à máquina reconhecer alguns comandos de voz em uma linguagem, “pode-se falar em um microfone conectado ao computador e o computador converte a fala em arquivos de textos ou comandos” (GOMES, p. 11, 2010). Logo, permite ao computador reconhecer a fala convertendo-a em texto. Essa conversão de fala em texto dialoga diretamente com o RV.

2.2 O QUE É RECONHECIMENTO DE VOZ?

O Reconhecimento de Voz é feito por um *software* do sistema que permite que a fala seja reconhecida e convertida em texto. De modo resumido, o RV funciona da seguinte forma: quando o/a usuário/a fala, a rede capta o que é proferido e codifica a fala, depois disso, por meio de um método determinado pela programação, a rede associa as informações ditas e, por meio de padrões de palavras, encontra uma

semelhança entre os termos e, assim, associa as palavras com seu padrão, alcançando o objetivo desejado pelo/a usuário/a, podendo ser desde o reconhecimento exato das palavras ditadas até ligar aparelhos eletrônicos e eletrodomésticos.

O RV faz parte de uma Inteligência Artificial. Como já mencionado, a IA foi criada para auxiliar, cada vez mais, o cotidiano de quem a utiliza. As Ferramentas das IAs, como Reconhecimento de Voz, – Google Assistente, Siri, Alexa, Cortana, Bixby, entre tantas outras – têm como objetivo facilitar a vida de quem as emprega, seja alternando entre os serviços de *streaming*, ligando para um contato, criando alarmes e *timers*, definindo lembretes ou até mesmo acendendo e apagando as lâmpadas dos cômodos de uma casa.

As Assistentes Virtuais, nome dado a essas ferramentas, têm como principal meio de funcionalidade a voz do/a usuário/a que a utiliza. É por meio de comandos feitos e compreendidos pela rede, apenas pela fala, que as Assistentes de Voz das IAs realizam o que é pedido – por meio do RV.

As IAs podem, então, serem definidas como sistemas de computadores que “pensam”, “agem”, de uma forma lógica. Por tal motivo, elas são ótimas ferramentas para realizar as tarefas mais comuns da rotina (como ler, escrever ou enviar mensagens; ligar a máquina de lavar roupa, a cafeteira, o ar-condicionado) de forma mais prática, agilizando a vida do usuário.

Por exemplo, uma pessoa que utiliza uma IA para enviar mensagens de texto, não precisa, necessariamente, digitar a mensagem, ela pode ditar para a IA, por meio do RV, e a IA transcreve o que é dito e envia o texto para o destinatário, após o comando. Desse modo, a pessoa consegue, ao mesmo tempo que dita, realizar outra tarefa.

2.3 A IARA HEALTH

Com o avanço da tecnologia, foi percebido que essa agilidade proporcionada pelo uso das ferramentas de RV poderia ser aplicada em diversas áreas, não somente no âmbito do lazer, mas também da educação, do comércio, do atendimento ao público e até da saúde. Tal uso não visa somente à praticidade, mas também propicia maior qualidade de vida ao usuário e melhor eficiência no trabalho a ser realizado.

Conforme dito anteriormente, o RV – tecnologia sobre a qual esse trabalho se deterá – pode ser integrado às Inteligências Artificiais. Essas tecnologias transformam a maneira de viver dos que as utilizam e tais mudanças podem ser, paulatinamente, mais perceptíveis tanto no Brasil quanto no exterior.

No cenário brasileiro, desde 2018, uma IA brasileira vem inovando e aprimorando a área da saúde por meio do Reconhecimento de Voz. A Iara, Assistente da *Iara Health*, é o primeiro sistema de RV 100% nacional que oferece maior agilidade, eficiência e qualidade de vida aos médicos/as radiologistas e aos pacientes.

A *Iara Health* tem um vocabulário especializado não somente no contexto, mas também no léxico, da Radiologia e utiliza a IA para garantir uma alta taxa de acerto na transcrição do que é dito pelo usuário (chamada de acurácia). Além disso, segundo o site¹ da *startup*, *Iara Health*, o produto é compatível com qualquer microfone e entrega a transcrição em tempo real, independentemente do sistema operacional que o usuário utilize – *Windows* ou *Mac OS*.

2.4 POR QUE OS RADIOLOGISTAS?

Os radiologistas são os profissionais da saúde que realizam a análise de exames de imagem. Tais exames podem ser ultrassons, tomografias, ressonâncias magnéticas, mamografias, dentre outros. Para realizar a análise, ou seja, gerar o laudo dos exames, esses profissionais precisam que um digitador² transcreva a sua análise ou necessitam que eles mesmos digitem seu parecer médico.

Com o surgimento da Iara, esse profissional não precisa se preocupar em digitar ou esperar o digitador, ele pode realizar seu laudo, sua análise, apenas ditando para a IA da *Iara Health* que transcreve em tempo real.

Por meio do seu sistema, a Iara permite que seus usuários – médicos e médicas radiologistas – acessem o portal *Iara Health* e ditem seus prontuários médicos. A Iara, então, garante que o profissional da radiologia transcreva em tempo real e de forma mais rápida os seus laudos. Dessa forma, a Iara melhora o trabalho dos radiologistas, tornando-o mais rápido e prático, permitindo que eles se preocupem com o diagnóstico e com os pacientes – uma vez que cada laudo médico se refere a um exame de um paciente, que necessita de atendimento preciso e de excelência.

A rede, especializada em Radiologia, tem uma taxa de 99% de acerto e pode melhorar esse percentual de acordo com o tanto que o usuário utiliza a rede, uma vez que ela é treinada para se adequar ao usuário.

¹ Disponível em: <https://iarahealth.com/>. Acesso em: 16 jan. 2023.

² Um digitador é um profissional que ouve a análise feita pelo radiologista e digita cada palavra dita por ele.

Contudo, foi observado que algumas palavras não estavam sendo reconhecidas de forma correta pela rede. Muitos são os fatores que podem estar influenciando para que essa transcrição não seja exata, desde sons do ambiente, problemas com o microfone, o tom de voz, o sotaque, a forma com que o usuário articula as palavras, suas variações linguísticas, entre outros.

A partir da constatação de que a ferramenta não tem 100% de acurácia, esse trabalho almeja investigar, analisar e entender – do ponto de vista linguístico e da fonética e fonologia – quais são os aspectos da linguagem que podem estar interferindo no reconhecimento correto das palavras ditadas pelos usuários da *Iara Health*.

Para isso, além dos textos escolhidos como *corpus* dessa pesquisa (apresentados na íntegra nos Anexos), foram usados como base e referencial teórico os textos de autores das áreas da Fonética e Fonologia e das áreas relacionadas à tecnologia que abrangem tanto Inteligência Artificial quanto RV e seus ramos.

Contudo, ressalta-se que, apesar de ser possível analisar e estudar esses mesmos pontos sob uma óptica da Computação, da Programação ou de outras áreas da tecnologia, esse trabalho se deterá em uma análise sob o viés da Linguística, mais precisamente, da Fonética e Fonologia, uma vez que esta área também está relacionada e fortemente interligada à Inteligência Artificial e ao Reconhecimento de Voz. Sendo assim, pode-se afirmar que essa área é tão essencial quanto as demais para o funcionamento de forma eficaz de uma IA, como ressaltou Gomes.

2.5 A SÍLABA NO PORTUGUÊS BRASILEIRO

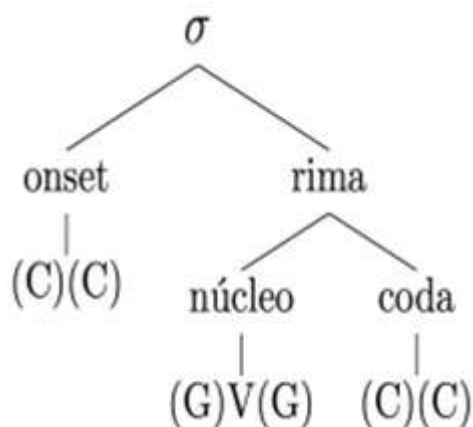
As sílabas são formadas por três elementos, quais sejam, vogais, consoantes e semivogais, ordenadas em combinações específicas.

As vogais são consideradas elementos obrigatórios para a existência da sílaba, elas ocupam a posição chamada “núcleo” (ou “pico silábico”). Já as consoantes e as semivogais ocupam as posições denominadas periféricas. É importante ressaltar que, ao contrário das vogais, as consoantes e as semivogais não são elementos obrigatórios para a existência da sílaba no Português brasileiro (doravante PB), logo ambas podem não ocupar nenhuma posição na sílaba.

As posições periféricas são chamadas assim porque abrangem as posições tanto anteriores quanto posteriores ao núcleo. Quando a posição ocupada é anterior ao núcleo, a posição pré-vocálica, tem-se o ataque ou *onset* silábico. Por outro lado, quando a posição ocupada é posterior ao núcleo, a posição pós-vocalica, tem-se a coda silábica.

Existem teorias fonológicas que entendem a sílaba como sendo uma construção hierárquica constituída por *Onset* e Rima. Um exemplo dessas teorias é a de Selkirk (1982), cuja estrutura interna da sílaba remete aos galhos de uma árvore. Nessa teoria, a sílaba – representada por “ σ ” – é ramificada em *onset* e em rima. A rima, então, se subdivide em núcleo e coda, como observado na Figura 1.

Figura 1- Estrutura da Sílaba do Português



Fonte: Adaptado de Bisol (1999).

As sílabas que existem no PB são divididas em quatro tipos: simples; complexas; abertas; e fechadas.

As sílabas simples são formadas apenas por núcleo, ou seja, apenas por vogais. Já as complexas são constituídas pelo núcleo e por uma consoante – independentemente se a consoante ocupa a posição pré ou pós vocálica. As sílabas abertas, por sua vez, são caracterizadas por não terem coda. Por último, as sílabas fechadas são aquelas que possuem coda.

2.5.1 Onset silábico

Como dito e exibido anteriormente, o *onset* silábico ocupa a posição pré-vocálica da sílaba – ou seja, anterior ao núcleo – e pode ser ocupado por até duas consoantes. Quando é ocupado por apenas uma consoante, o *onset* silábico é chamado de simples. Qualquer consoante pode ocupar essa posição em início de palavra, exceto

pelas consoantes /ɲ/, /ʎ/ e /r/ - exemplificando, respectivamente, *espinhal*, *ilhota* e *anexar* – uma vez que essas consoantes, no PB, só são usadas em meio de palavra.

Quando o *onset* é ocupado por duas consoantes, ele é chamado de complexo. Nesse caso, especificamente, a segunda consoante que ocupará a posição precisa ser uma consoante líquida, podendo ou ser uma lateral, /l/ - como em *active* – ou uma não-lateral, /r/ – como em *acrômio*. Na ocasião em que duas consoantes estão juntas, como é o caso das consoantes em *onset* complexo, dá-se o nome de Encontro Consonantal Tautossilábico (ECT). Porém os encontros “tl” e “vr” são encontrados apenas em meio de palavra, no PB – como pode-se observar em: *atleta*; e *neuralgia*. Já os encontros “dl” e “vl” não ocorrem nem mesmo no meio de palavra. É de suma importância lembrar que a Fonética e a Fonologia têm como objeto de estudo o som produzido e não a grafia, portanto a posição ocupada pelas consoantes citadas refere-se ao som e não à grafia.

2.5.2 A Coda Silábica

A coda ocupa a posição pós-vocálica³, ou seja, é posterior ao núcleo. Ela pode ser composta por uma ou por duas consoantes. É chamada de coda simples aquela que é composta por apenas uma consoante e de coda complexa a que é formada por duas consoantes. “Os segmentos fonéticos que ocorrem em coda silábica são normalmente representados por arquifonemas em função da variação que ocorre nessa posição. São eles: /N/, /S/ e /R/.” (SEARA; NUNES; VOLCÃO, 2011, p.97).

Um arquifonema é utilizado para representar que um ou mais fonemas perderam a distinção, logo, “um arquifonema expressa a perda de contraste fonêmico, ou seja, a neutralização – de um ou mais fonemas em um contexto específico” (SILVA, 2003, p.158).

A presença ou não do arquifonema /N/ varia de acordo com a teoria a ser seguida, uma vez que existem duas principais correntes que o tratam de formas distintas. A corrente de Câmara Jr. considera que o sistema fonológico do PB possui sete vogais orais – [a], [e], [ɛ], [i], [o], [ɔ] e [u]. A vogal nasal, por sua vez, seria construída por meio de um segmento vocálico oral com um segmento consonantal nasal – ou seja, a vogal nasal, para Câmara Jr. é feita de uma vogal oral acompanhada de uma

³ Em virtude das diferentes pronúncias encontradas no Brasil, a variante velar [ɰ] e a vocalizada [w] podem ser encontradas na posição pós-vocálica.

consoante nasal. Por ser constituída por dois fonemas, o autor a classifica como bifonêmica. Já a outra teoria de Head, Pontes e Back interpreta que asvogais nasais são fonemas, como pode observar a seguir:

[...] argumenta que o sistema fonológico do PB comporta sete vogais orais e cinco nasais. Ou seja, haveria palavras que se distinguem apenas pela nasalidade da vogal. Dessa maneira, ã e a seriam fonemas na língua, uma vez que formam pares mínimos nas palavras *cata* e *canta*, visto que o que as distingue são as vogais *a* (oral) e *ã* (nasal). Para tal corrente, não há necessidade do arquifonema /N/, e a palavra *canta* tem a seguinte transcrição fonológica: 'ãa. (SEARA; NUNES; VOLCÃO, 2011, p. 97).

O arquifonema /S/ é usado para evidenciar que o contraste dos fonemas, em coda silábica, /s z ʒ ʒ/ foi neutralizado. Diferentemente de quando estão em *onset* silábico, esses fonemas perdem o seu contraste quando estão na posição de coda silábica simples (como em *paz*) ou complexa (como em *rins*).

O destaque do arquifonema /R/ - que também pode aparecer em coda simples – é a sua característica de manifestar-se foneticamente de formas diferentes, no PB. Os sons correspondem a dois distintos fonemas no início de sílaba, os quais são chamados de fraco, /r/, e forte, /r/.

3 METODOLOGIA

Para a realização dessa pesquisa foram escolhidos textos e frases veículos para as palavras que continham diferentes moldes silábicos. Esses textos e frases se inserem em contexto médico, mais especificamente, o radiológico. Eles foram selecionados de forma aleatória, apenas contemplando a necessidade de apresentar pelo menos uma palavra da frase transcrita erroneamente pela rede Iara quando foram ditas por médicos radiologistas usuários da plataforma. Logo, pode-se inferir que as necessidades para selecionar os textos resumiram-se em: (i) não terem sido reconhecidos de forma correta quando proferidos; e (ii) possuir erros em diferentes segmentos silábicos.

Enfatiza-se que foi necessário obter tanto o texto falado quanto o texto transcrito, a fim de ser possível realizar uma comparação entre o texto original, falado, e o texto reconhecido, o resultado da transcrição feita pela rede, ou seja, o que realmente foi transcrito pela IA. Essa comparação foi feita observando as palavras reconhecidas, por meio de um levantamento de acertos e erros de cada uma delas.

Tendo como base esse compilado, as palavras reconhecidas erroneamente foram separadas das transcritas corretamente, voltando o olhar apenas para aquelas. Após a separação, foi feita uma classificação observando especialmente, os encontros consonantais e as codas silábicas – pois, em um primeiro momento, sem uma análise mais aprofundada, estes eram os principais impasses para o reconhecimento das palavras erradas na íntegra.

As categorias estabelecidas para tal classificação foram: C + r (e.g. *frontais*, *processo*, *parágrafo*); C + l (e.g. *aterosclerótica*, *active*, *fluida*); Coda Nasal (e.g. *hipoinsuflação*, *tentório*, *circunflexas*); Coda Fricativa (e.g. *recessos*, *obliterados*, *frontais*); Coda Lateral (e.g. *umbilical*, *realce*, *provavelmente*); Coda Rótica (e.g. *formações*, *cerebelar*, *vírgula*); Coda Ortográfica Epentética (e.g. “kt” - *aspectos*, *caracterizada*, *respectiva*); Coda Ortográfica Epentética (e.g. “gm” e “gn” - *segmentos*, *diafragma*, *significativas*); Coda Ortográfica Epentética (e.g. “bd” - *abdômen*, *abdominal*); Coda Epentética (e.g. “ks” - *tórax*, *anteversoflexão*).

Após a classificação das palavras, foi necessário separar as sílabas nas categorias citadas acima para melhor visualização e confirmação de estavam sendo atendidas. Essa separação, por sua vez, foi realizada em uma planilha, chamada de Tabela de Classificação Fonética, como pode ser observado no quadro a seguir, o qual apresenta as categorias estabelecidas e algumas palavras que compõe a tabela em questão:

Quadro 1 – Classificação fonética

Categorias	Palavras	
C+ r	frontais	processo
C+ l	aterosclerótica	active
Coda Nasal	hipoinsuflação	tentório
Coda Fricativa	recessos	realce
Coda Lateral	formações	cerebelar
Coda Rótica	cerebelar	germinoma
Coda Epentética - kt	aspectos	caracterizada
Coda Epentética - gm/gn	segmentos	diafragma
Coda Epentética - db	abdômen	abdominal
Coda Epentética - ks	tórax	anteversoflexão

Foi necessário trabalhar com as sílabas, ao invés de palavras inteiras, por se entender que as palavras são compostas por sílabas e essas são formadas por fonemas. Ou seja, os fonemas, quando pronunciados em conjunto formam sílabas que quando se unem a outras formam palavras, estas podem formar frases e textos.

Tendo finalizada a separação e confirmação de que as categorias estavam sendo respeitadas, foi feito um documento, em PDF, com todos os 33 textos e frases. Esse documento foi elaborado para enviar aos voluntários, a fim de que eles realizassem sua gravação no portal da *Iara Health* – disponível no Anexo A. Essa gravação foi necessária para que fosse possível ter o controle de que as palavras dos textos fossem ditas uma única vez por diferentes pessoas, usuários e dessa forma obter um resultado condizente com a realidade da rede da *Iara Health*.

As gravações foram feitas por cada uma das voluntárias, utilizando o Portal *Iara Health*. Elas foram aconselhadas a se familiarizar com os termos e gravar, de forma natural, cada um dos 33 textos. Em relação às voluntárias, é importante ressaltar que totalizaram três pessoas que se identificam com o gênero feminino. Todas já tinham alguma familiaridade com o Portal *Iara Health*, mas não eram médicas radiologistas, nem profissionais da saúde.

Elas trabalham na área de linguagem e processamento de dados da empresa e utilizam o portal para tarefas como testar recursos, verificar se algumas palavras estão sendo reconhecidas corretamente e outras possíveis funcionalidades que precisam ser experimentadas. Cabe ressaltar que apesar das voluntárias serem apenas mulheres, não foram controlados: o sexo; a idade; a escolaridade; e a região geográfica. A escolha das participantes para o teste dos textos se deu de modo espontâneo, partindo de cada uma quando foi mencionado que a pesquisa precisaria de voluntários.

Logo, as voluntárias receberam por e-mail, *WhatsApp* ou pela plataforma de comunicação *Discord* o arquivo em PDF contendo os 33 textos e as orientações para a realização da gravação. Tais orientações resumiram-se em informar às voluntárias que era necessário realizar a leitura de cada texto em voz alta, utilizando o Portal da *Iara Health*. Também foi solicitado às participantes que se familiarizassem com os textos, antes de realizar a gravação, isso foi necessário para evitar possíveis dificuldades de pronúncia que pudessem comprometer o entendimento da fala, uma vez que existiam termos médicos não muito conhecidos por pessoas de fora da área radiológica.

As orientações ainda informavam as voluntárias de que não se preocupassem com a forma como seria lido o texto em voz alta, pois não existe um jeito certo ou

errado de fazê-lo. As participantes, então, gravaram cada um dos 33 textos utilizando o Portal da *Iara Health* e, logo que terminaram as gravações, encaminharam os seus áudios, juntamente com os resultados obtidos. Por resultados obtidos entendem-se as transcrições feitas pela rede da *Iara Health*, após a leitura dos 33 textos.

Após as gravações e com as transcrições das voluntárias em mãos, foi feita, mais uma vez, a classificação das palavras transcritas incorretamente, segundo os resultados, as transcrições, de cada voluntária. Tanto a classificação quanto a análise, feita posteriormente a classificação, foram realizadas de forma individualizada, sem comparação entre as voluntárias. Destaca-se que para a realização da análise foi utilizada a Tabela de Classificação Fonética elaborada na fase inicial da pesquisa.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Após a separação, foi possível voltar o olhar para a estrutura interna da palavra. Ao observar essa estrutura, percebeu-se que os principais erros estavam relacionados com: *onset* silábico; similaridade fonética; outros; coda nasal; flexão de número; epêntese; não reconhecimento; hipersegmentação; ambiguidade estrutural; flexão de gênero; coda fricativa; coda nasal + *onset* complexo; flexão de número + flexão de gênero; prefixo; coda nasal; e sigla.

Destaca-se que as categorias de erros foram elencadas à medida que foram analisadas cada uma das palavras. Por esse motivo, algumas categorias novas passaram a compor a análise – como é o caso de “flexão de número” e “flexão de gênero”, por exemplo, como pode-se observar no quadro a seguir:

Quadro 2 –As Principais Categorias de Erros da Iara da *Iara Health*

Categorias de erros
Ambiguidade Estrutural
Coda Fricativa
Coda Nasal
Coda Nasal + Onset Completo
Epêntese
Flexão Gênero
Flexão Número
Hipersegmentação

Não Reconheceu
Onset Complexo
Outros
Prefixo
Similaridade Fonológica

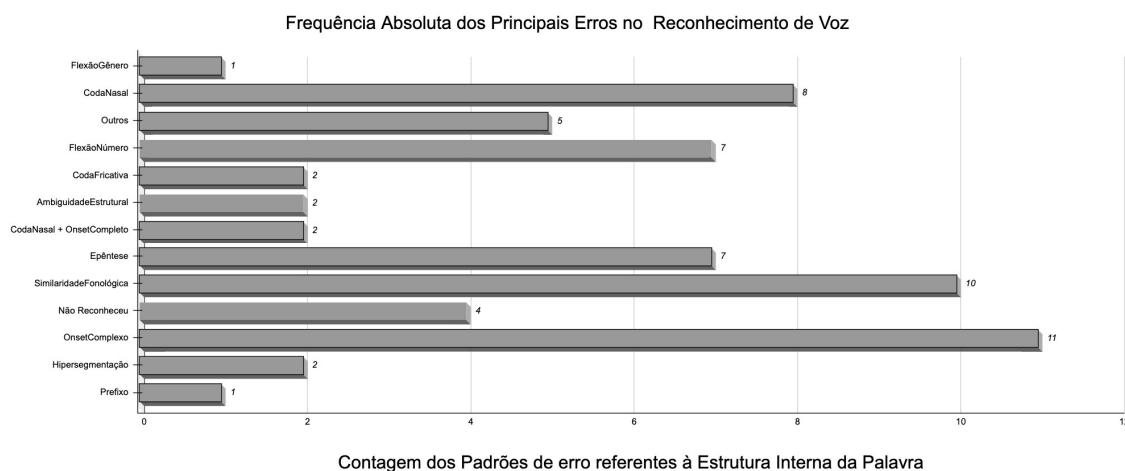
Após essa separação, foi possível realizar uma estimativa de palavras erradas e de palavras corretas – levando em consideração apenas o *corpus* deste trabalho. Os 33 textos continham, somados, um total de 509 palavras. Logo, foram contabilizadas a recorrência de cada um dos erros relacionados a cada uma das categorias distribuída para cada voluntária.

Ressalta-se que cada voluntária teve diferentes números de erros, a voluntária 1 apresentou 20 erros de todas as suas 509 palavras transcritas. Já a voluntária 2 teve apenas 12. A voluntária três foi a que apresentou mais erros, 41. Somando os erros, a rede da *Iara Health* teve um total de 73 erros. Isso significa que a rede em questão acertou, aproximadamente 96%, 97% e 91% das palavras para as três voluntárias, respectivamente.

4.1 RESULTADOS DA PERFORMANCE DA IARA

Entre as separações feitas, indica-se que os cinco principais tipos de erros da IA da *Iara Health* estão relacionados com: o *onset* complexo (verificando 11 ocorrências desses casos); a similaridade fonológica (apresentando 10 ocorrências); a coda nasal (contemplando 8 casos); a flexão de número e a epêntese (ambas com 7 casos cada); e os aspectos que não puderam ser contemplados nessa pesquisa, chamados de “outros” (agregando 5 casos). Como pode-se observar no Gráfico 1:

Gráfico 1 - Frequência Absoluta dos principais Erros no RV da Iara



Para exemplificação de *onset* complexo, no Quadro 3, estão as palavras *fluída* e *macroadenoma*, as quais foram reconhecidas como *e de sua* e *adenoma*, respectivamente.

Quadro 3 - Exemplos de Onset Complexo

Texto	Palavra desejada	Palavra reconhecida
considerar as hipóteses diagnósticas de craniofaringioma vírgula macroadenoma hipofisário invasivo e germinoma ponto	Macroadenoma	adenoma
Fluída	Fluída	e da sua

Para exemplificar a similaridade fonética, no Quadro 4, observam-se ocorrências da palavra *protusão*, que foi reconhecida como *ponto a*, e *hífen*, reconhecido como *em*.

Quadro 4 - Exemplos de Similaridade Fonética

Texto	Palavra desejada	Palavra reconhecida
-------	------------------	---------------------

prostrusão discal focal de base larga em situação posterolateral barra foraminal esquerda em l quatro l cinco vírgula indentando o saco dural e se insinuando no aspecto inferior do respectivo forame neural ponto parágrafo	prostrusão	ponto a
hífen quadrante superomedial na mama direita medindo zero vírgula nove centímetros ponto parágrafo hífen junção dos quadrantes superiores da mama esquerda medindo zero vírgula cinco centímetros ponto parágrafo	Hífen	Em

Para a coda nasal (contemplando 8 casos), são usadas como exemplo ocorrências das palavras *sendo*, a qual foi reconhecida como *sem*, e *parameninge que*, reconhecida como *parede lisa e*. Observe no Quadro 5.

Quadro 5 - Exemplos de Coda Nasal

Texto	Palavra desejada	Palavra reconhecida
desidratação do disco intervertebral de l cinco s um e parcial dos demais discos sendo mais evidentes em l três l quatro e l quatro l cinco ponto parágrafo	Sendo	Sem
parágrafo realce da parameninge que reveste as fissuras inter hemisférica e o tentório cerebelar ponto parágrafo	Parameninge	parede lisa e

Já para exemplificar a flexão de número, no Quadro 6, mostramos as palavras *seios*, reconhecida como *seio*, e *dos*, reconhecida como *do*.

Quadro 6 - Exemplos de Flexão de Número

Texto	Palavra desejada	Palavra reconhecida
conchas nasais médias bolhosas vírgula aeração do recesso lateral direito dos seios esfenoides	Dos	Do
conchas nasais médias bolhosas vírgula aeração do recesso lateral direito dos seios	Seios	Seio

O Quadro 7 traz exemplos da ocorrência de epêntese com as palavras *abdômen* e *subcondral*.

Quadro 7 - Exemplos de Epêntese

Texto	Palavra desejada	Palavra reconhecida
abdômen	abdômen	abdome
aspecto	aspecto	aspeto

Por último, o Quadro 8 exemplifica a categoria outros.

Quadro 8 - Exemplos de Outros

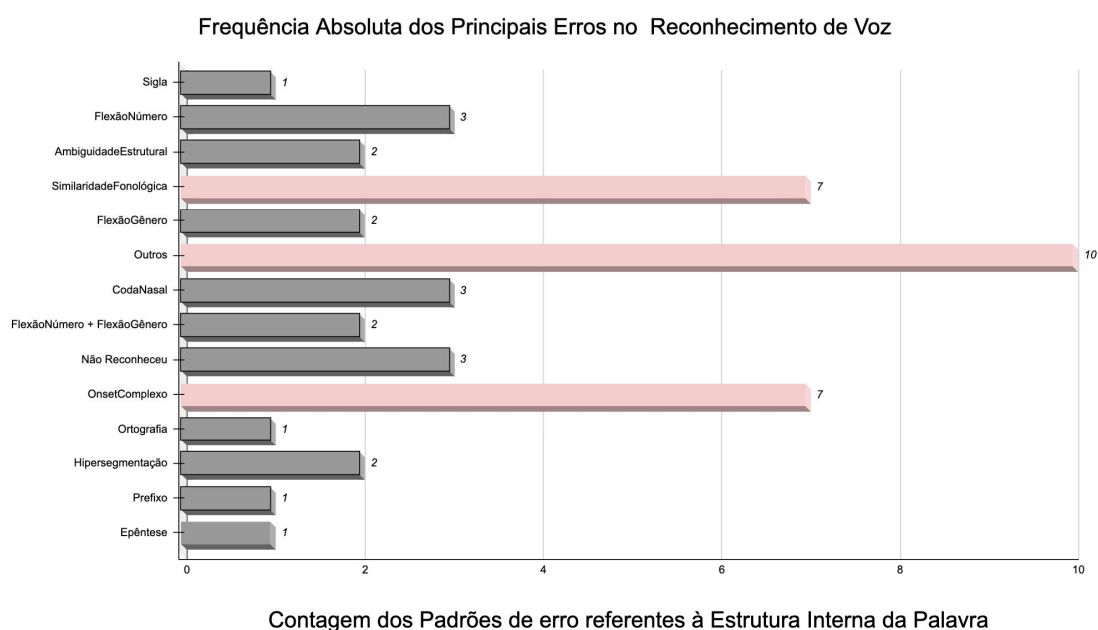
Texto	Palavra desejada	Palavra reconhecida
conchas nasais médias bolhosas vírgula aeração do recesso lateral direito dos seios esfenoides	nasais	evidências
com leve migração inferior vírgula tocando o saco dural bem como a raiz nervosa	nervosa	vírgula

4.2. COMPARAÇÃO IARA VERSUS GOOGLE ASSISTENTE

A título de comparação, os mesmos 33 textos foram ditos para a *Google* Assistente. Para realizar a comparação entre as redes, foi preciso separar cada palavra transcrita pela IA do *Google* da mesma maneira que foi feito com as palavras reconhecidas pela IA da *Iara Health*.

Após as separações feitas, foi verificado que os três principais tipos de erros da IA do *Google* estão relacionados com: outros (totalizando 10 ocorrências), seguido por similaridade fonológica e *onset* complexo (cada um deles contando 7 casos). O Gráfico 2 a seguir apresenta a frequência dos principais erros do RV na rede da *Google*.

Gráfico 2 - A Frequência Absoluta dos Principais Erros no RV da Assistente da Google



Em rosa, estão destacadas as três principais ocorrências de erros do RV da rede do *Google*: Outros apresentou 10 ocorrências; e Similaridade Fonológica e *Onset* Silábico apresentaram, cada, 7 ocorrências.

É importante ressaltar que as categorias que foram mais presentes no Reconhecimento da Assistente da *Google* não são as mesmas das que foram mais recorrentes no Reconhecimento da Iara. O Quadro 3 traz as principais categorias de erros presentes no teste feito utilizando a Assistente da *Google*.

Quadro 3 –As Principais Categorias de Erros no RV da *Google*Assistente.

Categorias de erros

Ambiguidade Estrutural
Coda Nasal
Epêntese
Flexão Gênero
Flexão Número
Flexão Número + Flexão Gênero
Hipersegmentação
Não Reconheceu
Onset Complexo
Ortografia
Outros
Prefixo
Similaridade Fonológica

As gravações para a Assistente da *Google* foram feitas apenas por uma das voluntárias, a voluntária três - a fim de verificar se a taxa de erros seria igual ou menor, já que, no teste realizado pelo Portal da *Iara Health*, ela foi a que mais apresentou erros. Da mesma forma foi levantada uma estimativa de acertos dessa rede - ainda considerando apenas as palavras dos 33 textos selecionados, totalizando 509. A Assistente da *Google* errou 143 palavras das 509 totais, sendo, aproximadamente, 28% de todo o *corpus* transcrito de forma incorreta.

5 CONSIDERAÇÕES FINAIS

Em virtude das análises ora mencionadas infere-se que os principais fatores linguísticos que comprometem o reconhecimento das palavras que compunham os 33 textos, o *corpus*, estão relacionados: ao *onset* silábico; à similaridade fonética; à coda nasal; à flexão de número; à epêntese; ao não reconhecimento; à hipersegmentação; à ambiguidade estrutural; à flexão de gênero; à coda fricativa; à coda nasal acompanhada de *onset* complexo; à flexão de número acompanhada da flexão de gênero; ao prefixo; à coda nasal; e à ocorrência de sigla.

Observou-se, também, que em relação ao reconhecimento, as palavras foram transcritas com mais acerto pela IA da *Iara Health* do que pela IA da *Google*, principalmente as palavras da área da Radiologia. Uma vez que a rede da *Iara Health* apresentou 41 erros (considerando apenas os resultados do teste realizado pela

voluntária três, já que esta foi quem obteve mais erros entre as participantes da pesquisa), enquanto a *Google* transcreveu 143 palavras de forma incorreta.

Recorda-se que a voluntária três, por ter apresentado uma alta taxa de erros utilizando a *Iara*, IA da *Iara Health*, foi a única que realizou a comparação com a IA da *Google*. Portanto, tem-se uma diferença de 102 palavras, aproximadamente 20% do total do *corpus*.

Considerando que o uso da rede é destinado para laudar prontuários médicos, a *Iara Health* demonstra ser a mais indicada para os radiologistas, já que tem uma taxa maior de acertos, em relação à ferramenta do *Google*, a qual tem um caráter mais genérico. O que significa que para o contexto da Radiologia, a *Iara* revela-se como a alternativa que acertou mais de 90% do *corpus* para todas as voluntárias.

Ainda que os fatores linguísticos apontados nessa pesquisa comprometam tanto o reconhecimento da *Iara Health* quanto da *Google* Assistente, foi observado que a IA da *Iara Health* conseguiu transcrever mais palavras corretamente do que a *Google*. Logo, pode-se inferir que, por ser uma solução específica para a área, a *Iara Health* apresenta uma precisão maior em suas transcrições do que as ferramentas genéricas de RV.

Cabe dizer, ainda, que apesar desta pesquisa levantar quais os principais fatores podem estar comprometendo o desempenho da IA da *startup* brasileira, a solução para esses obstáculos precisa ser desenvolvida em conjunto com as áreas de Desenvolvimento, Programação, Computação, entre outras que possam vir a ser necessárias. Isso comprova a multidisciplinaridade da Inteligência Artificial mencionada por Gomes (2010).

Para maiores detalhamentos ou até mesmo novas análises desses fatores linguísticos é recomendada a realização de uma pesquisa mais aprofundada do assunto, para que dessa maneira seja possível compreender, de forma ainda mais clara, a razão desses fatores estarem comprometendo o RV da Inteligência Artificial da *startup Iara Health*.

REFERÊNCIAS

BARROS, L. N.de. **Introdução à Inteligência Artificial Agentes Inteligentes.**

DocPlayer. Disponível em: <https://docplayer.com.br/26180733-Introducao-a->

inteligencia-artificial-agentes-inteligentes-leliane-nunes-de-barros.html. Acesso em: 16 jan. 2023.

BISOL, L. **Introdução a estudos de fonologia do português brasileiro**. 5ª edição, revisada. Porto Alegre: EDIPUCRS, 2010.

CAMARA Jr., J. M. **Estrutura da língua portuguesa**. Petrópolis: Vozes, 2001.

CARVALHO, A. C. P. de L. F. de. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados**, São Paulo, V. 35, n.101, p. 21-36, abr., 2021. Disponível em: <https://www.revistas.usp.br/eav/article/view/185020/171202>. Acesso em: 16 jan. 2023.

GOMES, D. Santos do. Inteligência Artificial: Conceitos e Aplicações. **Revista Olhar Científico**, Rondônia, V.01, n.2, p. X-X, ago., 2010. Disponível em: https://www.professores.uff.br/screspo/wp-content/uploads/sites/127/2017/09/ia_intro.pdf. Acesso em: 16 jan. 2023.

OLIVEIRA, D. H. de. **Fonética e Fonologia**. Disponível em: <https://www.studocu.com/pt-br/document/universidade-federal-do-tocantins/fonetica-e-fonologia/11dernerval-da-horo-fonetica-e-fonologia/40731779>. Acesso em: 20 jan. 2023.

SEARA, I. C.; NUNES, V. G.; LAZZAROTTO-VOLCÃO, C. **Fonética e Fonologia do português brasileiro**. Disponível em: http://ppglin.posgrad.ufsc.br/files/2013/04/Livro_Fonetica_e_Fonologia.pdf. Acesso em: 16 jan. 2023.

SELKIRK, E., *et al.* The syllable. **The structure of phonological representations**, v. 2, p. 337-383, 1982.

ANEXOS

ANEXO A – Texto enviado para as participantes

Prezado/a Participante,

A lista de texto abaixo constitui uma tarefa de leitura em voz alta a ser direcionada a uma Assistente Virtual. Peço que você se familiarize primeiramente, fazendo uma leitura silenciosa até se sentir confortável para realizar a gravação. Os textos são simples e curtos, mas talvez você encontre algumas palavras pouco familiares a você. Não se preocupe, você não está sendo avaliado quanto às pronúncias dessas palavras e não há certo ou errado. Apenas as leia como você faria para qualquer outro texto. Desde já, agradeço sua participação e disponibilidade para ajudar a pesquisa brasileira em Linguística. Quando se sentir preparado/a, leia cada texto conforme ele está escrito, até mesmo números e sinais de pontuação, da forma mais fluída e natural que você conseguir. Quando terminar de realizar as gravações dos 33 textos, por favor, me avise.

Muito obrigada por ter aceitado participar da minha pesquisa! Qualquer dúvida, pode me chamar tanto no Discord quanto por e-mail: saraapg009@gmail.com ou por Whats: (12)98706-6111.

Texto 1

recessos frontais obliterados por interposições de mucosa espessada ponto demais vias de drenagem pèrvias ponto parágrafo tc do dia dois do cinco de dois mil e vinte e dois ponto parágrafo

Texto 2

conchas nasais médias bolhosas vírgula aeração do recesso lateral direito dos seios esfenoides

Texto 3

cento e quarenta e nove vírgula sete

Texto 4

hérnia umbilical com insinuação de gordura vírgula exibindo alteração da densidade vírgula não se afastando processo inflamatório associado ponto parágrafo diástase da musculatura do reto abdominal ponto parágrafo demais aspectos descritos no corpo do relatório

Texto 5

baço vírgula pâncreas e vias biliares sem alterações ponto parágrafo formações nodulares de baixa densidade em ambas adrenais abre parênteses

Texto 6

multifocal ponto parágrafo as formações nodulares vascularizada no baço são inespecíficas vírgula porém mais provavelmente benignas ponto parágrafo demais aspectos avaliados descritos no corpo do relatório

Texto 7

fígado de morfologia habitual vírgula exibindo algumas formações císticas esparsas por ambos os seus lobos vírgula medindo até dois vírgula um centímetros nosegmentos dois ponto parágrafo

Texto 8

alguns pequenos achados esparsos em ambos os seus lobos vírgula mostrando abre aspas lavagem fecha aspas na fase tardia vírgula sem componente de hiper realce arterial ou mesmo abre aspas cápsula fecha aspas vírgula com dimensões subcentimétricas

Texto 9

fígado vírgula baço vírgula adrenais vírgula pâncreas e vias biliares sem alterações ponto parágrafo colecistectomia ponto parágrafo baço com morfologia preservada vírgula exibindo algumas formações nodulares vascularizadas no parênquima medindo até um vírgula três centímetros no terço superior

Texto 10

artérias renais pÉrvias vírgula sem áreas de estenose ou aneurismas vírgula identificando se ramo arterial para cada rim ponto parágrafo sistema arterial ilíaco mostrando opacificação preservada vírgula sem estenoses vírgula destacando se ectasia da ilíaca comum direita que mede até um vírgula sete centímetros de diâmetro ponto parágrafo

Texto 11

rins exibindo realce satisfatório pelo meio de contraste vírgula sem cálculos ou hidronefrose vírgula identificando se cistos simples bilaterais vírgula subcentimétricos ponto parágrafo formação nodular vascularizada na parede lateral esquerda do útero vírgula correspondendo provavelmente a mioma e medindo um vírgula nove centímetros ponto parágrafo

Texto 12

com leve migração inferior vírgula tocando o saco dural bem como a raiz nervosa

Texto 13

Labro

Texto 14

Labrum

Texto 15

desidratação do disco intervertebral de I cinco s um e parcial dos demais discos sendo mais evidentes em I três I quatro e I quatro I cinco ponto parágrafo

Texto 16

acrômio tipo dois de bigliani ponto parágrafo alterações degenerativas na articulação acromioclavicular vírgula caracterizada por irregularidade barra esclerose subcondral vírgula cistos e focos de edema nas superfícies

Texto 17

labro sem alterações significativas ponto parágrafo

Texto 18

protrusão discal focal de base larga em situação posterolateral barra foraminal esquerda em I quatro I cinco vírgula indentando o saco dural e se insinuando no aspecto inferior do respectivo forame neural ponto parágrafo

Texto 19

hífen quadrante *superomedial* na mama direita medindo zero vírgula nove centímetros ponto parágrafo hífen junção dos quadrantes superiores da mama esquerda medindo zero vírgula cinco centímetros ponto parágrafo

Texto 20

parágrafo realce da parameninge que reveste as fissuras inter hemisférica e o tentório cerebelar ponto parágrafo

Texto 21

considerar as hipóteses diagnósticas de craniofaringioma vírgula macroadenoma hipofisário invasivo e germinoma ponto

Texto 22

anteversoflexão

Texto 23

circunflexas

Texto 24

confluindo

Texto 25

deflexional

Texto 26

fluída

Texto 27

flutuante

Texto 28

hipoinsuflação

Texto 29

Active

Texto 30

Aterosclerótica

Texto 31

Aspecto

Texto 32

diafragma

Texto 33

abdômen