

# Sistemas de Diálogo Falado em Português: uma Abordagem Multimodal para a Saúde Digital

André Kazuo Takahata<sup>1</sup>, Antonio Valerio Netto<sup>2</sup>, Ricardo Suyama<sup>1</sup>,  
Margarethe Born Steinberger-Elias<sup>1</sup>

<sup>1</sup>CECS, UFABC, Av. dos Estados, 5001 - Bangú, Santo André - SP, 09210-580

<sup>2</sup>DIS, EPM/UNIFESP, Rua Botucatu, 862 - Têrreo, Vila Clementino, São Paulo- SP, 04023-062  
e-mail: andre.t@ufabc.edu.br

**Resumo:** Construir um sistema de diálogo falado em Português Brasileiro voltado a área da saúde é estratégico em um país com alto número de analfabetos funcionais. Este trabalho faz uma abordagem teórica inicial desse tema.

## 1. Introdução

Sistemas de diálogo (DS, *dialog systems*) consistem em tecnologias computacionais que simulam uma comunicação com um ser humano ou um personagem com características humanas, e assim, permitem a interação com o usuário por meio de linguagem natural, seja na forma de texto ou falada. Na literatura acadêmica e em empresas, os DS são chamados de agentes conversacionais, assistentes virtuais, sistemas humanos virtuais, *chatbots* ou *chatterbots* (texto) e *voicebots* (voz), entre outros. DSs podem, desde agendar compromissos por comando de voz, até facilitar a comunicação protocolar em domínios corporativo, médico, acadêmico, jurídico e outros. O volume crescente da demanda por dados digitais tem levado a inteligência artificial (IA) e o processamento de linguagem natural (PLN) a disponibilizar cada vez mais tecnologias interacionais de informação. É o caso dos assistentes pessoais lançados pelas grandes corporações de tecnologia como “Siri”, “OK Google” e “Cortana”.

Também na área da saúde cabe a demanda por *chatbots*, um exemplo é o brasileiro “Ana” do HC-UFMG<sup>1</sup>, que informa pacientes e faz pré-triagem de COVID-19. A literatura cita vários exemplos internacionais de sistemas de diálogo em saúde [1], seja no acesso a informações *online*, na promoção da saúde por meio do estilo de vida, seja na formação de profissionais de saúde, no monitoramento de pacientes crônicos, deficientes e idosos e também no auxílio à mudança de hábitos. Este tipo de sistema é escalável, podendo apresentar-se de forma ubíqua e pervasiva e acessível em dispositivos móveis e 24 horas por dia. Para a população brasileira, o impacto de um DS falado seria significativo, já que 29% são analfabetos funcionais entre 15 e 64 anos e 53% entre 50 e 64 anos, segundo o Indicador de Analfabetismo Funcional (INAF)<sup>2</sup>. Este trabalho apresenta características e desafios para pesquisa de sistemas de diálogo falado em português na área da saúde.

## 2. Algumas características dos sistemas de diálogo falados (SDS)

Segundo Deriu et al. [2] há três grandes classes de DS: a) orientados a tarefas, ajudam o usuário em fins específicos, tem uma estrutura derivada do seu domínio de conhecimento e um conjunto fechado de respostas possíveis; b) agentes conversacionais (*chatbots*), são diálogos menos estruturados em domínio aberto, tendem a ser mais longos e a emular interações sociais; c) sistemas de pergunta-resposta, respondem perguntas de domínio aberto ou fechado com base em amplos repositórios de dados (*data-driven*) e em modelos de linguagem que aprendem com as interações bem sucedidas. Os DSs distinguem-se, entre outros critérios, pelas modalidades de interação: escrita ou falada. Os sistemas de diálogo falados (SDS, *spoken dialog systems*) possuem os seguintes componentes: reconhecimento automático de voz (ASR, *automatic speech recognition*); entendimento de linguagem falada (NLU, *natural language understanding*); gerenciamento de diálogo (DM, *dialog manager*); geração de linguagem natural (NLG, *natural language generator*); síntese de voz (TTS, *text-to-speech*) [3]. O módulo ASR é responsável por converter o sinal de voz em dados textuais e extrair características de prosódia. Em seguida, a NLU obtém uma representação semântica por meio de tarefas como segmentação, identificação de intenções e reconhecimento de entidades nomeadas. O DM possui a tarefa de orquestrar a conversação, garantindo que haja uma troca de informações com o usuário de forma adequada para se atingir um objetivo. Para isso, baseado nas informações das entradas e saídas (presentes e passadas), bem como em uma base de conhecimento, o DM executa as tarefas de controle do estado da conversação (*dialog state tracking*), internamente ao sistema, e das decisões a serem tomadas. Assim, o DM gera comandos com representações

<sup>1</sup> <https://telessaude.hc.ufmg.br/>

<sup>2</sup> <https://ipm.org.br/inaf>

semânticas, que são convertidas em texto e marcas de prosódia no NLG. Essa saída, por sua vez, é convertida em um sinal de voz pelo TTS. É importante destacar que também existem sistemas que tratam de modo integrado um subconjunto de sequências (*sequence-to-sequence architecture*), ou até mesmo sistemas *end-to-end*, em que um único modelo é utilizado para formar o sistema.

DSs orientados a tarefas e agentes conversacionais (*chatbots*) tendem a ser multimodais, isto é, a utilizar informações de texto, voz e imagem. Geralmente tais fontes são tratadas em separado, ainda é um desafio avaliar como se dá sua integração e qual o impacto de cada modalidade no desempenho do agente. Por outro lado, a multimodalidade tem potencial na área médica, por exemplo, na transcrição de conversações médicas, na detecção de nível de estresse, nível de satisfação de serviço. Além disso, o uso de características de voz pode ser importante também para capturar mudanças de sentido pela modulação de entonação do usuário. Outro uso importante está no reconhecimento e expressão de sentimentos, visto que a adequação do discurso ao estado emocional do usuário é importante para seu engajamento.

### 3. Desafios para pesquisa no contexto do Português Brasileiro (PB)

O Português Brasileiro (PB) tem sido apontado como uma língua ainda com poucos recursos linguísticos de processamento (*under-resourced languages*), como corpora monolíngues, dicionários eletrônicos bilíngues, dados de fala transcritos, dicionários de pronúncia e glossários, se comparado ao inglês, francês, alemão ou espanhol. A pesquisa de SDS em português tem avançado devagar, embora o trabalho linguístico preliminar de pesquisa de uma gramática do português falado tenha dado largos passos. No PB, a ênfase tem sido o desenvolvimento de sistemas capazes de processar e sintetizar falas em português, com todas as suas variações da forma oral [4]. Morais et al. (2009) [5] criaram, a partir de uma interface de reconhecimento e síntese de voz da Microsoft, um protótipo de DS em português com entrada pelo componente ASR e saída pelo componente TTS, mas reconhecem que gramáticas livres de contexto e “aplicações para ditado continuam distantes do desejável diálogo espontâneo”. Além disso, é importante que o sistema seja capaz de identificar dificuldades do usuário no seu entendimento de conceitos da área da saúde e adequar a saída para que o conteúdo possua uma linguagem acessível a todos os usuários [6].

Com respeito às particularidades na Saúde, deve-se levar em consideração a segurança e a privacidade do usuário, em especial a dados considerados sensíveis pela Lei Geral de Proteção aos Dados (LGPD). Também é importante estabelecer critérios de avaliação de desempenho técnico, experiência do usuário e sobre a real eficácia e eficiência do seu uso na área da saúde [1]. Outro ponto importante na elaboração de sistemas de diálogo está no combate a vieses algorítmicos, que consiste em corrigir erros sistemáticos que possam discriminar ou privilegiar um grupo de pacientes.

### 4. Considerações finais

Embora haja sistemas de diálogo falado em português disponibilizados pelas grandes empresas de tecnologia na forma de assistentes digitais de uso geral, existem ainda diversos desafios para viabilizar seu uso pleno na área da saúde. Atualmente, existe uma escassez de diretrizes nacionais e recursos que possam assegurar a elaboração de sistemas robustos para apoiar a ações de Atenção à Saúde em nosso país.

### Referências

- [1] LARANJO, Liliana et al. “Conversational agents in healthcare: a systematic review”. *Journal of the American Medical Informatics Association*, v. 25, n. 9, p. 1248-1258, 2018.
- [2] DERIU, Jan et al. “Survey on Evaluation Methods for Dialog Systems”. arXiv:1905.04071v2[cs.CL]26 Jun 2020.
- [3] YOUNG, “Steve J. Probabilistic methods in spoken–dialogue systems”. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, v. 358, n. 1769, p. 1389-1402, 2000.
- [4] LIMA, Thales Aguiar de; COSTA-ABREU, Márjory da. “A survey on automatic speech recognition systems for Portuguese language and its variations”. *Computer Speech & Language*, v. 62, p. 101055, 2020.
- [5] MORAIS, Jefferson; NETO, Nelson; KLAUTAU, Aldebaro. “Tecnologias para o desenvolvimento de sistemas de diálogo falado em Português Brasileiro”. In: *7th Brazilian Symposium in Information and Human Language Technology*. 2009.
- [6] FINATTO, M.J.B. “Acessibilidade textual e terminológica: promovendo a tradução intralinguística”. *Revista Estudos Linguísticos*, v. 49, n. 1, p. 72-96, abr. 2020.