



Conversando com máquinas: construindo um chatterbot especializado em Literatura Francesa para aulas de Francês Língua Estrangeira

Paulo Bruno Lopes da Silva

(GREat-UFC)

Mardônio Jó de França

(GREat-UFC)

Resumo

Este trabalho utiliza um chatterbot, ferramenta computacional interativa que simula conversações entre homem e máquina, no processo de formação de aprendizes de Francês Língua Estrangeira (FLE), além de análise qualitativa dos resultados dessa interação, visando contribuir para a análise da conversação, área que sistematiza interações verbais e diálogos, e para o desenvolvimento de chatterbots. O artigo objetiva apresentar etapas de criação da base de diálogos de um chatterbot especializado em literatura francesa e verificar sua contribuição para o ensino de línguas. Para isso, o estudo concentra-se no conceito de Corpus Especial (SINCLAIR, 1996) e nos princípios e métodos da Análise da Conversação de Kebrat-Orecchioni (2006). A construção do chatterbot foi estruturada nas etapas de Configuração do Perfil, Mineração dos Dados, Geração da Base de Diálogos, Aprendizagem de Máquina e Interação Homem-Máquina. Nas duas primeiras, foram escolhidas obras de 10 escritores franceses, gerando aproximadamente 2000 registros posteriormente usados no modelo de aprendizagem de máquina. Em seguida, essa aplicação será avaliada por aprendizes de FLE, que interagem e classificam as respostas segundo sua aceitabilidade, validando a última etapa. Assim, esperamos formar um chatterbot capaz de manter conversas com alta aceitabilidade, apresentando-se como nova ferramenta para o processo de ensino de línguas estrangeiras.

Palavras-chave: Chatterbot, Análise da Conversação, Aprendizagem de Máquina.

Abstract

This work uses a chatterbot, interactive software tool that simulates conversations between man and machine, in the French apprentice training process as Foreign Language, as well as improving qualitative analysis of the results of this interaction, to contribute to the Conversation Analysis, area that systematizes verbal interactions and dialogs, developing chatterbots. This article presents the steps of creating the base of a specialized



chatterbot in French literature and verify its contribution to language teaching. Then, this study focuses on the concept of Special Corpus (Sinclair, 1996) and the principles and methods of Conversation Analysis proposed by Kebrat-Orecchioni (2006). The construction of the chatterbot was structured in steps: Profile Configuration, Data Mining, Dialogues Base Generation, Machine Learning and Human Computer Interaction. In the first two, 10 French writers and works were chosen, generating about 2000 records later used in machine learning model. Then this application will be evaluated by learners of FLE, interacting and classifying answers according to their acceptability, validating the last step. Thus, we expect form a chatterbot able to hold talks with high acceptability, presenting itself as a new tool for foreign language teaching process.

Keywords: Chatterbot, Conversation Analysis, Machine Learning.

Introdução

Desde os primórdios da História, a tecnologia tem sido um diferencial de desenvolvimento destacado na sociedade. Aplicada a diversos contextos, surgem diferentes ferramentas e aplicações com o intuito de facilitar atividades e tarefas do cotidiano.

Com o desenvolvimento de tecnologias como os computadores, os dispositivos móveis e, sobretudo, da internet, a possibilidade de utilização de recursos educacionais por meio de aplicações e ferramentas tem sido bastante estimulada.

Em se tratando do ensino e aprendizagem de Francês Língua Estrangeira (FLE), um problema bastante evidente é gerado pela dificuldade que os aprendizes enfrentam para praticar a língua-alvo fora do contexto de sala de aula. Seja por falta de acompanhamento de professor seja pela deficiência na comunicação com outros alunos e aprendizes ou a dificuldade de encontrar falantes nativos e situações nas quais devam utilizar a língua estrangeira.

Para resolver tal situação, surgem diferentes aplicações que visem incentivar o uso de línguas estrangeiras na comunicação, como os softwares de comunicação instantânea (Skype, Google Hangouts etc.) e as redes sociais (Facebook, Google+ etc.).



Uma outra alternativa se encontra na utilização de chatterbots, agentes conversacionais virtuais, capazes de simular falantes por meio do Processamento de Linguagem Natural (PLN).

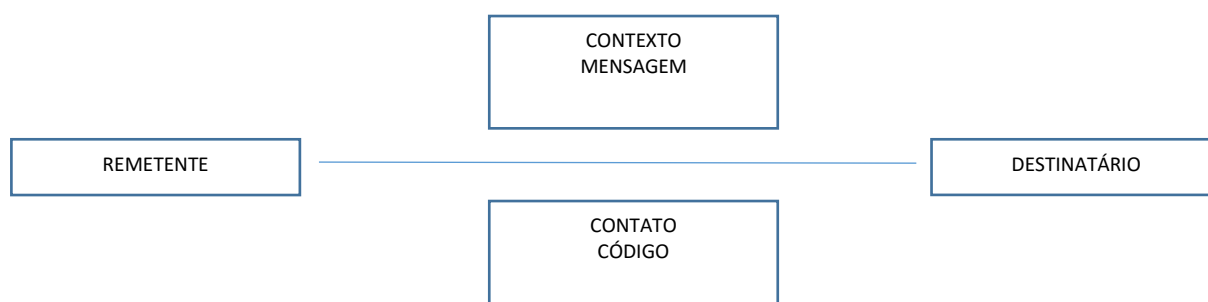
Portanto, o objetivo deste trabalho é apresentar uma proposta de metodologia de criação de um chatterbot especializado em Literatura Francesa com o intuito de ajudar no processo de ensino e aprendizagem de Francês Língua Estrangeira.

Para isso, nas seções a seguir, descrevemos as etapas de criação com viés linguístico, baseado nos princípios de interação verbal e Análise da Conversação propostos por Kebrat-Orecchioni (2006), bem como a análise do chat como novo gênero textual e sua transmutação na web (ARAÚJO, 2004).

1. Comunicação, interação e conversação

Nos moldes da Linguística, vários modelos foram propostos para explicar o processo de comunicação, porém o mais conhecido é o proposto por Jakobson (1969):

Figura 1: Modelo de Comunicação de Jakobson (1969)



Fonte: Jakobson apud Barros, 2007.



Segundo Barros (2007), o modelo proposto por Jakobson presume, no processo de comunicação, os seguintes elementos:

- Um remetente que envia uma mensagem a um destinatário;
- Um contexto (ou um referente) ao qual se refere a mensagem e que seja compreensível pelo destinatário;
- Um código, total ou parcialmente comum a ambos;
- Um contato, isto é, um canal físico e uma conexão psicológica entre o remetente e o destinatário, que os capacitem a estabelecer e manter a comunicação.

Tais elementos acabam por caracterizar os pontos básicos para se estabelecer uma comunicação entre dois participantes (emissor e receptor). No entanto, o processo de comunicação evolui em nível de complexidade, com a alternância de papéis entre os participantes.

De acordo com Kebrat-Orecchioni (2006, p. 27),

Em uma conversação, os participantes são chamados a ocupar, cada um a seu turno, a posição emissora (ficando, no resto do tempo, relegados à posição receptora), ou seja, as configurações interlocutivas não cessam de se modificar ao longo do desenrolar da interação.

Assim, toda troca comunicativa mostra-se mais complexa do que uma simples transmissão de uma mensagem de um emissor para um receptor por meio de um canal. Na verdade, é necessário que haja um “engajamento” por parte dos participantes para que se estabeleçam verdadeiras interações verbais e sociais.



Por fim, a conversação se estabelece como um tipo particular, e ao mesmo tempo prototípico, das interações verbais. Ressalta-se também que para que haja uma conversação, as interações verbais estão regidas por regras conversacionais (KEBRAT-ORECCHIONI, 2006, p. 14) que podem, ou não, estar associadas a um ou outro gênero em particular.

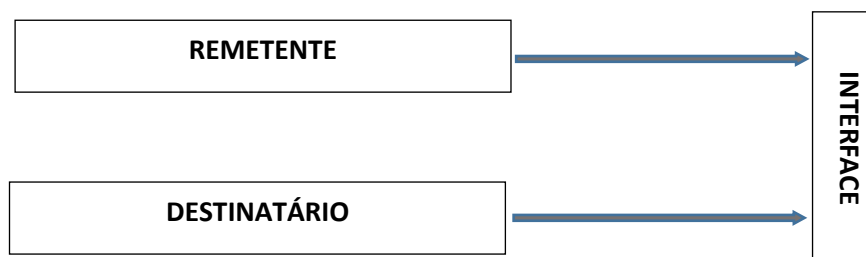
2. O Gênero *Chat*

Com o advento das novas tecnologias e as constantes mudanças nos meios de comunicação, surgem novos gêneros textuais característicos. Esses gêneros emergentes são relativamente variados, mas a maioria deles tem similares em outros ambientes, tanto na oralidade como na escrita (MARCUSCHI, 2002).

Nesse contexto, o chat surge como um gênero textual emergente dentro dos ambientes virtuais e, portanto, dotado de características próprias que o assemelham e o diferenciam de outros gêneros textuais existentes.

O chat ou bate-papo é um gênero textual eletrônico semelhante ao e-mail funcionando de forma síncrona, isto é, destinatário e remetente estão conectados simultaneamente na rede, o que transforma chats em conversas (ou bate-papo) em tempo real entre dois ou mais participantes (AMARAL; AMARAL, 2008).

Figura 2: Modelo de Comunicação no gênero *Chat*



Fonte: Elaborada pelos autores.



A existência do chat como gênero textual acontece, como já mencionado, pela semelhança e pela diferença que estabelece com outros gêneros existentes. O chat, aproxima-se das características apresentadas pela conversação espontânea face a face, porém diferencia-se por seu caráter de anonimato ou distanciamento físico entre os participantes.

Dessa forma, Marcuschi (2002) afirma que são criadas “novas formas de organizar os relacionamentos interpessoais nesse novo enquadre participativo”.

3. Chatterbot

Chatterbot são softwares que tentam emular a conversação humana por meio da demonstração de comportamentos semelhantes aos de uma pessoa sobre um domínio limitado. O principal uso dessas aplicações é o entretenimento, entretanto, chatterbots são empregados em atendimento a clientes, ferramentas de comunicação instantânea, jogos e até mesmo na educação (ABU SHAWAR; ATWELL, 2007).

O objetivo do funcionamento do robô é fazer com que este responda a determinados solicitações de entrada, os inputs (saudações, perguntas, pedidos), retornando saídas, os outputs, solicitados.

Como definem Ferreira e Uchôa (2006, p. 24), apesar de ter um propósito simples de definir, a implementação de um programa de computador capaz de atingir esse propósito é algo extremamente complexo. Sistemas de conversação automática como os chatterbots sofrem a limitação de não possuírem um aprendizado tão rápido e eficiente quanto o natural.

Na literatura, há três gerações de Chatterbots. Cada geração usa técnicas e implementações distintas. A primeira implementação usa padrões de similaridade e regras gramaticais, representada pelo software ELIZA, a segunda usa modelos de redes neurais para gerar respostas aos padrões de input, implementadas no software Julia. A



última implementação usa diferentes abordagens, tais como o AIML (Artificial Intelligence Markup Language), tendo como exemplo o chatterbot ALICE, e o Chat Script, um sistema para manipulação de linguagem natural.

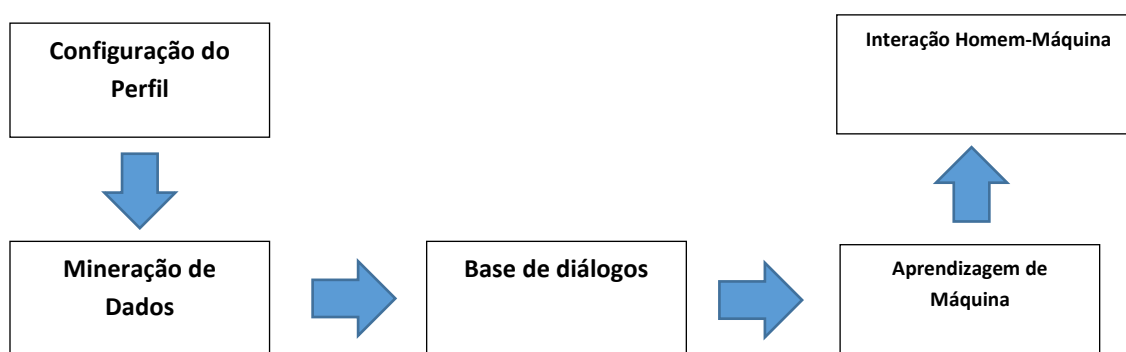
4. Metodologia Esfera

A construção de um chattebot é um processo complexo, pois envolve um grande número de processos, variações de domínios de aplicação, bem como problemas de processamento de linguagem natural.

A fim de facilitar a construção desse tipo de aplicações, esse artigo propõe a criação de uma nova metodologia de criação de chatterbots utilizando Aprendizagem de Máquina. A essa nova metodologia demos o nome de Esfera.

A Metodologia Esfera é composta por seis fases distintas: Configuração do Perfil, Mineração de Dados, Construção da Base de diálogos, Aprendizagem de Máquina e Interação Homem-Máquina. Essas etapas são mostradas na figura 3.

Figura 2: Etapas da Metodologia Esfera



Fonte: Elaborada pelos autores.



4.1 Configuração do perfil

A Configuração do perfil consiste em constituir a identidade do chatterbot, levantando pontos quando à sua personalidade e domínio. Esses elementos são a base para a criação da futura base de conhecimento do robô, pois, como é proposto por Franklin e Graesser (1996) o robô deve possuir um domínio (temas para o diálogo) e autoconhecimento (capacidade de o robô falar sobre si mesmo).

Assim, por meio da Metodologia Esfera, criamos um documento com os principais elementos característicos para criar o modelo de Identidade do robô Charles, como mostrado na tabela 1.

Tabela 1: Perfil do Chatterbot Charles

| DADO | MODELO DE IDENTIDADE |
|---------------|--|
| Nome | Charles |
| Gênero | Masculino |
| Nacionalidade | Francófono |
| Língua | Francês |
| O que faz | Gosta de estudar e conversar sobre Literatura Francesa |
| Especialidade | Especialista em 10 autores e suas respectivas obras maiores. |

Fonte: Elaborada pelos autores.

4.2 Mineração de Dados

A etapa seguinte consiste na Mineração de Dados. Durante a definição do perfil do robô Charles, foram selecionados dez autores da Literatura Francesa e suas respectivas obras maiores:

- Molière – L'avare;
- Honoré de Balzac – Le Père Goriot;
- Guy de Maupassant – La parure;
- Albert Camus – L'étranger;



- Simone de Beauvoir – Le deuxième sexe;
- Jean-Paul Sartre – Huis Clos;
- Victor Hugo – Les Misérables;
- Proust – À la recherche du temps perdu;
- Baudelaire – Les fleurs du mal;
- J. M. G. Le Clézio – Ritournelle de la faim.

O processo de mineração de dados pode ser efetuado de três formas: automática, semiautomática ou manual. Inicialmente, a coleta e seleção de materiais foi realizada de forma manual. No entanto, o processo automatizado tende a ser uma etapa por vir. Esses dados vão ser modelados para a etapa seguinte da metodologia.

4.3 Criação da Base de Diálogos

Após a seleção de textos proporcionada pela etapa de mineração, os dados foram modelados para que fossem transformados em estruturas conversacionais (inputs e outputs) que formariam a base de conhecimentos do robô Charles.

Para a etapa de criação da base de diálogos, aplicamos a metodologia proposta por Cunha e Silva (2015) para a geração automática de corpus de sentenças por meio de gramáticas livres de contexto. Nessa metodologia, utilizamos o NLTK (Natural Language Toolkit), biblioteca de ferramentas para o processamento de linguagem natural na linguagem Python (BIRD et al. 2009).

As entradas e saídas foram geradas simulando possíveis estruturas de conversação e, posteriormente pareadas, associando inputs aos respectivos outputs produzidos pelo chatterbot (Tabela 2).



Tabela 2: Pareamento de entradas e saídas da base de diálogo

| <i>INPUT</i> | <i>OUTPUT</i> |
|-----------------------------|---|
| Bonjour | Bonjour |
| Bonsoir | Bonsoir |
| Bonne nuit | Bonne nuit |
| Salut | Salut |
| Ça va ? | Ça va |
| Vous allez bien ? | Je vais bien, merci. |
| Comment allez-vous ? | Je vais bien, merci. |
| Comment vous vous appelez ? | Je m'appelle Charles. |
| Vous parlez français ? | Oui, je parle français et un peu de portugais. |
| Vous êtes Français ? | Je suis un robot francophone ! |
| Comment tu t'appelles ? | Je m'appelle Charles. |
| Tu parles français ? | Oui, je parle français et un peu de portugais. |
| Tu es Français ? | Je suis un robot francophone ! |
| Quelle est ta profession ? | J'aime parler et discuter sur la littérature française. |

Fonte: Elaborada pelos autores.

A criação desse corpus de sentenças automáticas também é amparada pelo conceito de corpus especial proposto por Sinclair (1996) um corpus pode ser produzido para fins específicos e que não têm a intenção de contribuir para a descrição exaustiva de fenômenos gramaticais.

4.4 Aprendizagem de Máquina

Após a geração e o pareamento de estruturas de entrada e saída das possíveis interações verbais, a Metodologia Esfera o uso de ferramentas de Aprendizagem de Máquina para gerar padrões e o modelo a ser utilizado para o processamento do chatterbot. Nesse projeto, utilizou-se o WEKA (Waikato Environment for Knowledge Analysis, Framework amplamente utilizado pela comunidade científica em problemas de inteligência artificial (FRANÇA et al., 2014).



Figura 4: Aprendizagem de Máquina na Metodologia Esfera



Fonte: Elaborada pelos autores.

4.5 Interação Homem-Máquina

Por último, após gerado o modelo de predição de respostas, foi possível passar para a etapa de execução ou o momento de interação homem-máquina, na qual pode-se analisar as relações de conversação entre emissor e receptor por meio do gênero chat (Figura 5).

Figura 5: Interação Homem-Máquina



Fonte: Elaborada pelos autores.



5. Resultados e Discussão

Após a aplicação das etapas adotadas pela Metodologia Esfera, obteve-se a primeira versão do chatterbot especializado em Literatura Francesa. Por meio da geração automática de sentenças foi possível gerar uma base de dados de 4529 interações verbais em língua estrangeira, envolvendo 4529 sentenças de input e 45 tipos de output.

Essa disparidade entre entradas e saídas é resultado das variações gramaticais das estruturas de pergunta do Francês, para as quais existem até três estruturas diferentes para sua formalização.

Além disso, mostra-se necessária, em uma etapa posterior, a geração de mais sentenças de resposta a fim de que o modelo seja aperfeiçoado.

Outra causa levantada para essa diferença pôde ser percebida ainda durante a etapa de criação do perfil: a polidez. Por questão de abrangência, foram geradas no corpus, sentenças formais (estruturas com vous) e informais (estruturas com tu). Em alguns casos, o número de combinações de estruturas de entrada dobrou quando pareadas às saídas desejadas.

Outro ponto relevante analisado na construção da base de diálogos equivale à estrutura inicial de inputs e outputs na forma de perguntas, aproximando o chatterbot, nesse momento, a um outro tipo de sistema, chamado de Q&A (Question and Answer). Esse modelo acaba sintetizando de forma primitiva o esquema de comunicação de Jakobson (1969) e afastando-o da estrutura de conversação de Kerbrat-Orecchioni (2006).



Considerações finais

O Chatterbot Charles, especializado em Literatura Francesa é um chattebot experimental criado para testar estratégias envolvendo abordagens de Aprendizagem de Máquina e PLN.

Foi possível perceber que a utilização de um chattebot para o ensino de Línguas Estrangeiras pode exercer um grande diferencial impactante sobre as novas metodologias de educação a distância por meio da tecnologia.

Como trabalhos futuros, percebe-se, portanto, a necessidade de ampliação da base de diálogos e dos pares de entradas e saídas, bem como, sobretudo, a avaliação dos outputs através de testes com aprendizes de Francês Língua Estrangeira.

Esses testes são fontes também para a implementação de estruturas que aproximem ainda mais os chatterbots de um sistema de conversação e ensino.

Referências Bibliográficas

ABU SHAWAR, B.; ATWELL, E. (2007). Chatbots: are they really useful? LDV-Forum – Band 22(1), 31-50.

AMARAL, Luiz H.; AMARAL, Carmem L. C. Tecnologias de comunicação e comunicação aplicadas à educação. In: Interações virtuais: perspectivas para o ensino de Língua Portuguesa a distância, São Carlos, 2008. p.11-20.

BARROS, Diana Luz Pessoa de. A comunicação humana. In: FIORIN, José Luiz. **Introdução a Linguística: objetos teóricos**. São Paulo: Contexto, 2007. p. 25-52.

CAVALCANTE, K. B. Chatterbot em Língua Portuguesa usando Aprendizagem de Máquina. XXXIII Encontro de Iniciação Científica, UFC, Fortaleza, 2015.

CUNHA, T. M.; SILVA, P. B. L. A criação de um corpus de sentenças através de gramáticas livres de contexto. IV Jornada de Descrição do Português, UFRN, Natal, 2015.

FERREIRA, Leandro Padilha; UCHÔA, Joaquim Quintero. **Desenvolvimento de um chatbot para auxiliar o ensino de Espanhol como Língua Estrangeira**. 2006. Disponível em: <[http://repositorio.ufla.br/bitstream/1/9629/1/ARTIGO_Desenvolvimento de um chatbot para auxiliar o ensino de espanhol como língua estrangeira.pdf](http://repositorio.ufla.br/bitstream/1/9629/1/ARTIGO_Desenvolvimento%20de%20um%20chatbot%20para%20auxiliar%20o%20ensino%20de%20espanhol%20como%20língua%20estrangeira.pdf)>. Acesso em: 31 out. 2015.



FRANKLIN, S; A GRAESSER,. **Is it an Agent, or just a program?:** A taxonomy for Autonomous Agents. 1996. Disponível em: <<http://www.msci.memphis.edu/~franklin/AgentProg.html>>. Acesso em: 13 nov. 2015.

Galvão, A.M; Barros, F.A; Neves, A.M.M; Ramalho, G.L. (2003). Persona-AIML: uma arquitetura para desenvolver chatterbots com personalidade. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Pernambuco, Recife, Pernambuco.

KERBRAT-ORECCHIONI, C. *Análise da conversação: princípios e métodos*. São Paulo: Parábola Editorial, 2006.

MARCUSCHI, L. A. Gêneros textuais emergentes no contexto da tecnologia digital. In: MARCUSCHI, L. A.; XAVIER, A. C. *Hipertexto e gêneros digitais*. Rio de Janeiro: Editora Lucerna, 2004. p. 13-67.

MAULDIN, M. L. **Chatterbots, TinyMUDs, and the Turing Test:** Entering the Loebner Prize Competition, Proceedings of Twelfth National Conference on Artificial Intelligence. Seattle: Aai-94, 1994

MITKOV, R. *The Oxford Handbook of Computational Linguistics*. University Express, 2009.

Weizenbaum, J. (1966). ELIZA A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 10(8):3645.

Witten, J.H ; Frank, E.; Hall, M.A. (2011). *Data Mining - Practical Machine Learning Tools and Techniques*. Elsevier. USA

Harris, Zellig (1954). *Distributional Structure*. *Word* 10 (2/3): 146-62.1.