

## A Utilização de Atos de Diálogo em Sistemas de Diálogo para Dispositivos Móveis

Tiago Martins da Cunha<sup>1</sup>, Daniel de França Brasil Soares<sup>2</sup>

<sup>1</sup>Instituto de Humanidades e Letras – Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

<sup>2</sup>Grupos de Redes de Computadores Engenharia de Software e Sistemas (GREat)

tiagotmc@unilab.edu.br, danielsoares@great.ufc.br

**Abstract.** *This work presents an approach towards the categorization of turn into a conversation between human and machine. We present the strategies for storing and recalling data into a chatterbot system architecture. All dialog turn classification and their theoretical principle to be selected into the system are, in this paper, presented. The use of such classification showed satisfactory conversation quality according to human evaluators. This evaluation was conducted taking into account the maxims sorted out by [Grice et al. 1975]. Our satisfactory results in comparison to other systems that propose chatting modules in a mobile system still require improvement and has great potential to become a noted conversational agent system.*

**Resumo.** *Este trabalho apresenta uma abordagem para a categorização de turnos em uma conversa entre o homem e máquina. Nós apresentamos as estratégias para armazenar e acessar os dados para a arquitetura de um sistema chatterbot. Todas classificações de turnos em diálogos e seus princípios teóricos, a ser utilizado na classificação, serão apresentados. O uso de tal classificação mostrou qualidade de conversação satisfatória de acordo com os avaliadores humanos. Esta avaliação foi realizada tendo em conta as máximas elencadas por [Grice et al. 1975]. Nossos resultados satisfatórios em comparação com outros sistemas, que propõem módulos de conversação em um sistema móvel, ainda precisam de melhorias e tem um grande potencial para se tornar um notável sistema agente de conversação.*

### 1. Introdução

O universo de aplicações do conhecimento sobre o Processamento de Linguagem Natural (PLN) é muito diversificada. Dessa forma, as aplicações de PLN podem usar uma vasta variedade de recursos linguísticos em suas arquiteturas. Neste trabalho, pretendemos apresentar uma abordagem que usa recursos oriundos da teoria da pragmática linguística para a modelagem discursiva de um sistema de diálogo para uma aplicação móvel. Bem como discutir as questões teóricas sobre a seleção dos atos de diálogo mais pertinentes para a nossa aplicação em um recurso móvel.

Sistemas de diálogo são conjuntos de processos que visam a manutenção de uma conversação. O nosso sistema é um chatterbot, que simula, em um telefone celular, um

agente conversacional, i.e. um indivíduo. Sistemas, desse tipo, foram muito populares na década de 80, e voltaram a ficar em evidência devido ao grande desenvolvimento de novas abordagens para manipulação de dados linguísticos e a melhoria do desempenho de processamento das máquinas.

A melhoria de desempenho permite diferentes métodos para análise e geração textual. Um desses métodos recentemente utilizados em sistemas de diálogos como chatbots é o modelo baseado em exemplos. Os modelos baseados em exemplos apresentam amostras de turnos em diálogos. A peculiaridade do nosso trabalho está no enriquecimento de classificação e distribuição de cada turno da conversação em Atos de Diálogo (DAs). O uso de DAs aumenta a espessura da base de conhecimento linguístico e auxilia na análise dos dados de entrada de um elocutor (usuário).

DAs agrupam as informações linguísticas pertinentes especialmente à pragmática, as quais classificam as sentenças utilizadas em cada turno de um diálogo. A teoria da DAs origina-se junto aos estudos pragmáticos dos Atos Illocutórios e da sua possível aplicação no diálogo entre homem e máquina.

Este tipo de informação linguística na base de dados permite-nos, em nossa arquitetura, analisar e prever a interação conversacional. A previsão depende de uma medida de similaridade matemática entre a entrada e o banco de dados. Através desta medida de similaridade, a saída do sistema é selecionada e fornecida ao utilizador. Cada entrada fornecida pelo usuário é analisada linguisticamente e verificada a sua similaridade com os dados previamente cadastrados no banco. A análise linguística da entrada sugere a sua relação com uma DA. No sistema, as DAs agrupam as possíveis intenções discursivas do usuário.

Cada sentença de entrada no banco de dados das DAs corresponde a uma saída como resposta do sistema. A fluidez na conversação que o sistema aspira depende da combinação entre entrada e saída. A satisfação nessa combinação é um dos parâmetros para medição da qualidade da conversação com nosso sistema. A avaliação de uma conversa não é uma tarefa fácil. Para resolver esse problema, nós projetamos uma avaliação para os sistemas de diálogo e a comparação do nosso sistema com outros disponíveis no mercado para a plataformas móveis.

Para esta avaliação, 9 voluntários foram solicitados para criticar a qualidade das respostas. Nesta avaliação nosso sistema obteve a melhor qualidade na interação com o usuário em comparação com outros sistemas que usam esse recurso. O parâmetro de avaliação era a satisfação do usuário em relação às respostas do sistema a serem consideradas a interação positiva, negativa ou neutra.

Nessa avaliação o nosso sistema obteve os maiores escores, mas mesmo obtendo a melhor interação com o usuário, existe ainda uma ampla variedade de estudos para serem feitos para melhorar ainda mais o nosso sistema. Conhecimentos Linguísticos, como de análise sintática e análise semântica lexical, podem melhorar o ranking de similaridade entre entrada e os dados contidos no banco para uma melhor atribuição de uma entrada a uma DA.

## 2. Atos Illocutórios

As regras da comunicação são adquiridas ao logo do desenvolvimento socio-cognitivo humano e são elas as responsáveis por governar o desempenho dos usuários de uma certa linguagem. A interação humana, em linguagem natural, segue inúmeros parâmetros estabelecidos através de um contrato social. Os falantes aprendem a se comunicar verbalmente através de regras linguísticas, *stricto sensu*, e de regras interacionistas, *lato sensu*.

Chamamos de regras linguísticas as que estruturam a língua e que passam por nossa linha de processamento, *pipeline*, e.g. regras morfolossintáticas, sintáticas, semânticas e pragmáticas. Sobre esse processamento, há ainda o campo da *praxis*, da interação. As regras interacionistas focalizam o contexto de fala, o momento da cooperação dos falantes no ato da conversação e serão principalmente essas as regras que fundamentam nosso trabalho. Sabe-se, pois, que o ato comunicativo não está livre totalmente a depender do fluxo do diálogo [Grice et al. 1975].

Nesse fluxo de diálogo, os interlocutores são capazes de identificar os objetivos da enunciação, os atos ilocucionários. Dentro da vasta classificação de intenções ilocucionárias propostas na literatura, nos propusemos a utilizar a classificação de atos ilocutórios expressivos [Searle 1969], e.g. pedir desculpa, agradecer, dar as boas vindas, etc. Esse tipo de classificação norteia a tipologia de domínios em nosso sistema de chatterbot citados na introdução deste trabalho, no mapeamento dos turnos de conversação.

Para se definir um ato ilocutório, faz-se necessário identificar as "condições de felicidade", as quais consistem em objetivo ilocutório, estado psicológico e conteúdo proposicional [Searle 1979]. Essas condições são tratadas em nosso trabalho para validar o uso das sentenças satisfatórias nos turnos de diálogo para a sua manutenção, por isso as nomeamos de condições de validação.

Tendo em vista que os atos ilocutórios têm a capacidade de representar a força ilocucionária do falante a que nosso sistema será modelado, poderíamos supor *a priori* que esse sistema para um chatterbot teoricamente seria capaz de classificar padrões de desejo, gratidão, desculpas, etc, no entanto, ainda precisaríamos tratar do conteúdo proposicional que não se insere dentro desses atos.

### 2.1. Atos De Diálogo

A fim de estruturar nosso sistema de diálogos faz-se ainda necessário mapear fragmentos menores da conversação que estão inseridos na sequência dialogal, mas que não contribuem para o significado principal do ato ilocutório. Para modelar uma conversação com esse tipo de estrutura, servimo-nos da classificação dos atos de diálogo [Stolcke et al. 2000].

Nesse trabalho, fundamentamo-nos, principalmente, no pressuposto de que é possível sistematizar o uso da linguagem. Os atos de diálogo, assim como os atos de fala, tentam descrever de forma sistemática os fenômenos pragmáticos da linguagem em contextos de fala específicos, no nosso caso, em ambiente virtual de interação entre homem e máquina.

Os DAs são responsáveis por classificar e modelar automaticamente estruturas do discurso, neste trabalho *inputs* de usuários do sistema de chatterbot. Essa função

pragmática segue a mesma orientação da teoria dos atos de fala no nível da força ilocucionária que representa o significado de proposições em contextos de uso real da linguagem [Austin 1975].

Para reconhecer os turnos de conversação dentro de nosso sistema, nos fundamentamos principalmente no trabalho de [Stolcke et al. 2000] o qual apresenta 48 tipos de DAs mapeados a partir do corpus Switchboard, um corpus oral de conversação por telefone [Godfrey et al. 1992].

### 3. Metodologia

O nascimento de Modelos baseados em exemplo são oriundos no universo dos estudos de Tradução que atribuiu essa abordagem como um tipo de Tradução Automática. De acordo com [Kay 1997], o par texto original e texto traduzido é determinado através da interface humana durante o processo de tradução. Essa mesma analogia foi levantada por [Nagao 1984] e pode ser utilizada para sistemas de diálogo.

Dessa forma, propomos uso de Modelos de Diálogo Baseado em Exemplos (EBDM, i.e. *Example-Based Dialog Models*) de acordo com [Lee et al. 2009] como abordagem para o nosso chatterbot. Em nosso sistema, o EBDM possibilita resultados interativos.

O EBDM é composto por exemplo de sentenças prototípicas para cada DA em que está agrupado. que serão comparados com a entrada do usuário exatamente (um jogo completo) ou através de uma medida de similaridade. Todas as frases de exemplo tem uma saída correspondente. Por exemplo: "Oi, como vai você?" está relacionada com a saída "Estou bem, obrigado". Além disso, o sistema procura a frase exata de entrada no banco de dados (memória). Se ele já está na base de conhecimento, a entrada é conhecida, dessa forma, já apresenta uma resposta adequada emparelhada a ela. No caso da entrada exata não estar armazenada no banco de memória, a correspondência mais provável será utilizada para proporcionar uma resposta. A melhoria do sistema depende da interação e colaboração humana na adequação das respostas do sistema.

Um sistema que utiliza EBDM tem de organizar os dados forma criteriosa. Na próxima seção, vamos apresentar abordagens para lidar com os dados através de um agente de conversação ou chatterbot.

#### 3.1. Escolha de DAs

O momento inicial da criação da nossa arquitetura do sistema dependeu da convenção sobre que tipo de interação esperaríamos que o usuário tivesse com o nosso sistema. Nesse momento propusemos 27 tipos de interação que foram classificadas posteriormente em DAs.

Como os tipos de interação só levaram em consideração o texto de entrada do usuário, elencamos 27 DAs para as entradas. Em seguida, produzimos um corpus de sentenças prototípicas para cada DA. Na tabela abaixo, apresentamos uma pequena amostra de sentenças contidas nos DAs.

Quadro 1: Amostra de DAs e entradas

Saudação	olá, tudo bem?
	Fala aí, Lígia!
Conselho	tu poderia não ser agressivo
	you deveria ser fiel
Terceira-Pessoa	a lua está muito bela
	que salada ruim
Insulto	you é uma lesma
	te acho uma vaca manca
Opinião-pergunta	you curte futebol?
	o que you pensa sobre a independência do Brasil?

### 3.2. O sistema

O sistema que propomos apresentava todos os dados organizados na plataforma SQL. Nessa plataforma cada sentença de entrada apresentava uma ou mais correspondentes de saída. Dessa forma, as sentenças de entrada eram agrupadas de acordo com o seu sentido, e.g. apresentavam as mesmas saídas. Assim por diante, cada grupo de sentenças de entrada convergiam para um DA, e este DA apresentava possíveis saídas.

Uma pequena amostra da arquitetura da organização dos dados pode ser vista na tabela abaixo. Na tabela podemos observar quatro exemplos de entrada que podem ser ditas pelo usuário, suas saídas, seus agrupamentos e em que DA estão classificados.

Tabela 1. SQL de entrada e saída

DA	Grupo	Entrada	Saída
Saudação	id_001	Oi!	Oi!
		Olá!	Olá!
	id_002	Como vai?	Tudo bem!
		Tudo bem?	Tudo!

Essa amostra apresenta o alinhamento entre entradas e saídas. Em nosso sistema, uma vez que a entrada exata do usuário está prevista pelo sistema, a resposta do sistema é oferecida. No entanto, se a resposta exata não está previamente cadastrada, é utilizado a medida de similaridade, e.g. medida Cosseno. Esta medida de similaridade usa como base os agrupamentos e prevê a semelhança do texto do usuário com cada agrupamento. É do grupo que mais se assemelha ao texto de entrada do usuário que a opção de saída será oferecida pelo sistema.

Caso o índice de similaridade seja muito baixo em relação aos agrupamentos, é escolhido o DA que obteve maior índice. A saída alinhada a um DA é uma resposta padrão e é reconhecida como um escape do sistema para manter sua robustez na manutenção do diálogo. No entanto, quando o sistema faz uso das respostas de escape, as entradas dos usuários são armazenadas para tratamento futuro e solucionar o problema de classificação da sentença de entrada individual e agrupada.

Esse tratamento é feito de maneira semi-supervisionada, da mesma forma que os agrupamentos. Esse processo foi feito periodicamente à medida que o banco de dados de

entrada ia crescendo devido ao uso. Diferentemente da Tradução Automática que fornece uma confirmação da qualidade da resposta do sistema ao editá-la, o tipo de interação, como em nosso sistema, não permite essa verificação. Dessa forma, precisamos passar por rotineiras etapas de avaliação.

### 3.3. A avaliação

O processo de avaliação da qualidade da interação por nosso sistema depende da avaliação humana. Procuramos realizar a avaliação como sugerido pela literatura de avaliação em sistemas de tradução automática[Koehn 2009].

No entanto, percebemos que os parâmetros de avaliação automática elencada pela teoria não atenderia o nosso propósito. Assim, a metodologia de avaliação foi reformulada para que houvesse uma validação satisfatória desse tipo de interação.

Selecionamos amostras dos pares de entrada e saída de diferentes DAs. As amostras de entrada foram submetidas a outros sistemas que fazem o uso de chatterbots e suas saídas foram coletadas.

Uma vez coletadas todas as entradas e saídas dos sistemas, os pares de entrada e saída foram avaliados por 9 sujeitos. Estes sujeitos julgaram se a saída do sistema era uma interação satisfatória com amostras o texto de entrada distribuídas entre todas DAs do sistema. Os sujeitos atribuíram nota entre um e três, sendo um para interação equivocada, dois para uma interação neutra, i.e. sem propriedades negativas ou positivas para a interação e três para completa relação da saída com o texto de entrada.

## 4. Discussão

Os resultados da avaliação são apresentados na tabela abaixo. A avaliação mostra que o nosso sistema, nomeado Lígia, obteve os melhores resultados em comparação com os outros sistemas.

Chamamos a atenção que os demais sistemas só disponibilizam seus recursos de chat conectados à internet. Já o nosso recurso foi utilizado *offline* (i.e sem conexão com a internet). Dessa forma, nosso sistema apresentava um banco de dados e processamento limitado.

**Tabela 2. Avaliação comparativa de qualidade de saídas**

Lígia	Voicemate	SIRI	Svoice
2.89	2.67	2.53	1.98

A obtenção desse resultado positivo, de escore 2,89, reforça a utilização de tratamento linguístico ao texto de entrada para classificar um diálogo e também indica a possibilidade de sistematização de elementos pragmáticos que se apoie em parâmetros linguísticos conversacionais.

Durante o tratamento linguístico, enfatizamos duas etapas basilares para que o processamento automático seja otimizado posteriormente. O primeiro passo consiste em obter um corpus representativo, que contenha os fenômenos linguísticos concernentes ao domínio-alvo do sistema. Em se tratando de um sistema de conversação, é necessário que tenhamos bastante abrangência das sequências de diálogo, como scripts de conversação.

Além de um corpus representativo, é importante também que haja dentro do sistema agrupamentos de sentenças que façam parte de um mesmo DA, a fim de que seja elaborado um grupo de sentenças de saída.

## 5. Conclusão

Nossa pesquisa elencou 27 DAs para classificar possíveis entradas do usuário. Acreditamos que elencar mais DAs pode possibilitar uma melhor classificação da interação entre homem e máquina. Ressaltamos que elencar DAs que não sejam ambíguas ou estruturalmente conflitantes não é uma tarefa trivial.

Nosso sistema, como prova de conceito para a classificação de interações através de DAs, mostrou-se satisfatório, mas precisamos ressaltar que ainda há muito o que fazer para melhorar o tratamento dos dados no sistema.

A própria atribuição de DAs ao texto de saída pode repercutir na qualidade do sistema. O texto de saída classificado em DAs permitiria um pareamento de texto de entrada e saída, tanto de sentença por sentença ou grupo por sentença quanto também DA de entrada por DA de saída.

Uma das grandes melhorias seria a implementação da geração de respostas através da classificação do texto de entrada do usuário. Essa geração poderia ser utilizada no momento em que a classificação submetesse a entrada ao texto de saída de escape. Esse recurso aumentaria a robustez do sistema.

Outra importante melhoria dar-se-ia ao agregar informações de cunho semântico ao processamento. Informações semânticas, como as contidas na WordNet e ConceptNet, favoreceriam o processo de classificação de DAs.

Acreditamos que a execução de uma pesquisa desse porte e a implementação de um sistema como esses só pode vir a agregar conhecimento e ampliar a aplicação de conhecimentos oriundos da linguística e suas aplicações computacionais.

## Referências

- Austin, J. L. (1975). *How to do things with words*, volume 367. Oxford university press.
- Godfrey, J. J., Holliman, E. C., and McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, volume 1, pages 517–520. IEEE.
- Grice, H. P., Cole, P., and Morgan, J. L. (1975). Syntax and semantics. *Logic and conversation*, 3:41–58.
- Kay, M. (1997). The proper place of men and machines in language translation. *machine translation*, 12(1-2):3–23.
- Koehn, P. (2009). *Statistical machine translation*. Cambridge University Press.
- Lee, C., Jung, S., Kim, S., and Lee, G. G. (2009). Example-based dialog modeling for practical multi-domain dialog system. *Speech Communication*, 51(5):466–484.
- Nagao, M. (1984). A framework of a mechanical translation between japanese and english by analogy principle. *Artificial and human intelligence*, pages 351–354.

- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge university press.
- Searle, J. R. (1979). *Expression and meaning: studies in the theory of speech arts*. Cambridge University Press.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., and Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.