DOI: 10.14393/DL38-v13n2a2019-4

A intenção discursiva nos sistemas de interpretação automática: um estudo bibliográfico

Discoursive intention in machine interpreting systems: a bibliographic study

Flávio de Sousa Freitas[*]
Marileide Dias Esqueda[**]

RESUMO: A interpretação automática é uma tecnologia que realiza a tradução de trechos de fala de uma língua para outra através da integração de três tecnologias, a saber, reconhecimento automático de fala, tradução automática e síntese de voz (FREITAS; ESQUEDA, 2017). Para desenvolver sistemas desse tipo, vários pesquisadores, primeiramente, buscaram compreender a fala humana, suas características, como ela é processada pelo cérebro humano e produzida pelo aparelho fonador (LEE, 2015). Somente após tais investigações é que foi possível o desenvolvimento de técnicas e abordagens para processar, representar e reproduzir a fala humana através de sistemas computacionais. Diante das inconsistências dos primeiros resultados, que representavam interpretações muito literais e errôneas, os estudiosos argumentam a favor de uma abordagem que leve em consideração a intenção do falante, aludindo ao contexto e aos propósitos comunicativos dos diálogos (JEKAT; KLEIN, 1996). Assim, este estudo bibliográfico busca explorar as características gerais e a evolução da abordagem de captura da intenção

ABSTRACT: Machine Interpreting or Speech-to-Speech Translation is a new technology that converts spoken utterances from one language into another through three functionalities grouped in only one software: Automatic Speech Recognition, Machine Translation and Speech Synthesis (or Text-To-Speech) (FREITAS; ESQUEDA, 2017). In order to develop systems of this type, several researchers first sought to understand human speech, its characteristics, how it is processed by the human brain and produced by the speech apparatus (LEE, 2015). It was only after such investigations that it was possible to develop techniques and approaches to process, represent and reproduce human speech through computational systems. Given the inconsistencies of the first results, which represented very literal and erroneous interpretations, scholars argue in favor of an approach that takes into account the speaker's intention, alluding to the context and the communicative purposes of the dialogues (JEKAT; KLEIN, 1996). Thus, this bibliographic study seeks to explore the general characteristics and evolution of

[*] Mestrando do Instituto de Letras e Linguística, Universidade Federal de Uberlândia. flaviofreitas@ufu.br
[**] Doutora em Estudos da Tradução e Professora Associada do Instituto de Letras e Linguística, Universidade Federal de Uberlândia. marileide.esqueda@ufu.br

discursiva em softwares de interpretação automática.

the discursive intention capture approach in machine interpreting software.

PALAVRAS-CHAVE: Interpretação Automática. Sistemas de Interpretação Automática. Intenção discursiva.

KEYWORDS: Machine Interpreting. Machine Interpreting Systems. Speaker's Intention.

1. Introdução

A Interpretação, como atividade humana de traduzir oralmente um discurso entre duas línguas diferentes, existe há muito tempo. Como atividade profissional, ela não é tão antiga assim. Suas primeiras aparições ocorreram no final da Primeira Guerra Mundial, nas conferências de Paris, na França, resultando no Tratado de Versalhes e no estabelecimento da Liga das Nações, atualmente Organização das Nações Unidas (PAGURA, 2010). A primeira escola a formar intérpretes foi fundada em Genebra, em 1941, por Antonie Velleman, um intérprete da Liga das Nações (BOWEN; BOWEN, 1984). E muitos outros cursos de formação de intérpretes foram surgindo ao longo das décadas seguintes.

Os primeiros intérpretes eram autodidatas, porém, com o avanço e importância das relações internacionais entre os mais diversos países, os cursos de formação em interpretação passaram a buscar procedimentos pedagógicos que pudessem preparar os futuros profissionais. As indagações que norteavam tal iniciativa diziam respeito à busca por interpretações de qualidade e, ao mesmo tempo, produzidas de forma instantânea e precisa.

A despeito de seus tipos, interpretação de conferências, em cenários empresariais, de tribunal, comunitária, ou de suas modalidades, consecutiva, simultânea, intermitente ou sussurrada, apenas para citar algumas, os Estudos da Interpretação como área acadêmica (PÖCHHAKER, 2004) sempre refletiram sobre quais fatores poderiam melhor contribuir para uma interpretação rápida, eficiente e livre de erros. Ao mesmo tempo em que tais fatores passaram a ser investigados com relação às interpretações humanas, o mesmo ocorreu com as pesquisas sobre a automação do trabalho dos intérpretes (JEKAT; KLEIN, 1996).

A união entre pesquisadores dos campos das Ciências da Computação, da Linguagem e do Processamento de Língua Natural e Artificial vêm se esforçando para implementar sistemas de interpretação automática, ou, em inglês, *speech-to-speech machine translation.* As primeiras publicações sobre o tema surgem na década de 1980, data também dos primeiros protótipos, que recebiam fundos milionários para levar adiante o que chamavam de *machine interpreting,* ou, em português, interpretação automática (doravante IA).

A IA é uma tecnologia que realiza a tradução de trechos de fala de uma língua para outra através da integração de três tecnologias, a saber, reconhecimento automático de fala (em inglês, *Automatic Speech Recognition* - ASR), tradução automática (em inglês, *Machine Translation* – MT; ou TA, em português) e síntese de voz (em inglês, *Speech Synthesis* ou *Text-To-Speech* - TTS)[1] (FREITAS; ESQUEDA, 2017).

Para desenvolver sistemas desse tipo foi necessário, primeiramente, compreender a fala humana[2] e suas características. Em seguida, investigou-se como ela é processada pelo cérebro humano e produzida pelo aparelho fonador. Somente depois disso foi possível o desenvolvimento de técnicas e abordagens para processar, representar e reproduzir a fala humana através de sistemas computacionais.

Diante das inconsistências dos primeiros resultados, que representavam interpretações muito literais e errôneas, Jekat e Klein (1996), e também vários outros pesquisadores da mesma década, argumentam a favor de uma abordagem que leve em consideração a "intenção discursiva", aludindo ao contexto e aos propósitos comunicativos dos diálogos, ou, para as autoras, em inglês, *intended interpretation*.

---

[1] Ao longo do trabalho, serão utilizados os acrônimos ASR, em referência ao reconhecimento automático de fala; TA em referência à tradução automática, sendo este um acrônimo já amplamente utilizado em publicações brasileiras; e TTS em referência à síntese de voz.

[2] A expressão fala humana empregada neste estudo refere-se ao ato comunicativo espontâneo, sem interferência de controladores e restrições linguísticas, tal qual se dá no dia a dia dos falantes de determinada língua.

Assim, com base em trabalhos anteriormente produzidos pelos autores deste artigo, que buscou investigar a arquitetura da tecnologia de IA, seus conceitos, definições e componentes, este novo estudo busca explorar as características gerais e a evolução da abordagem de captura[3] da intenção discursiva em softwares de IA. Além desta introdução, este artigo consiste em três seções, nas quais se busca retomar os achados das pesquisas envolvendo a IA, a partir dos dados de Freitas (2016) e Freitas e Esqueda (2017), descrever os aspectos metodológicos que regem esta nova busca, explorar brevemente como os autores aqui investigados definem a intenção no processamento de fala humana e mostrar os resultados das pesquisas envolvendo as formas de captura da intenção discursiva em sistemas de IA.

Faz-se importante destacar que não é propósito deste trabalho valorizar a automação em interpretação em detrimento do trabalho humano dos intérpretes. Em vez disso, vislumbra-se averiguar os avanços tecnológicos da interpretação automática, tema ainda pouco investigado no Brasil e no exterior, e que é, desde sua concepção, espelhada na interpretação realizada por humanos.

Como forma de justificar o estudo que ora se apresenta, busca-se contextualizar duas propostas antecedentes, os trabalhos de Freitas (*op. cit.*) e de Freitas e Esqueda (*op. cit.*).

2. Antecedentes teóricos sobre a IA

[3] Para o presente estudo optou-se pelo emprego do termo em português "captura", em referência ao termo em inglês *capture,* devido à sua maior frequência nos artigos aqui investigados. Outro termo que pode ser encontrado com menor frequência é "extrair" (ou *extract,* em inglês), também utilizado na literatura consultada. Como poderá ser constatado no decorrer do trabalho, a captura do significado de uma frase em linguagem oral, pronunciada por humanos, pode ser realizada por sistemas computacionais baseados em estatística ou em regras de estruturas gramaticais, com o propósito de ser traduzido para outras línguas também em linguagem oral. O maior interesse dos estudiosos dedicados à captura da intensão discursiva em sistemas de interpretação automática centra-se na questão de como tais sistemas podem capturar informações implícitas contidas na fala humana.

O estudo implementado por Freitas (*op. cit.)* teve como meta determinar o estado da arte da IA, a fim de investigar seus conceitos e suas definições, ainda ausentes nas teorizações pertencentes aos Estudos da Interpretação no Brasil. Neste trabalho, o autor realiza uma revisão bibliográfica a partir da década de 1980 até os dias atuais, tomando como base artigos da imprensa escrita e publicações em geral como aquelas presentes no Google Acadêmico, IEEEXplore, interACT, ACM-DL, *Interpreting: International Journal of Research and Practice in Interpreting* e em publicações dos laboratórios ATR Spoken Language Translation Research Laboratories.

Os objetivos foram coletar os conceitos concernentes à IA, embora ainda não totalmente sedimentados, buscando identificar os principais teóricos que cunharam os termos a ela relacionados, em sua maioria em cenário internacional, elaborar um quadro resumitivo com a cronologia das principais definições propostas pelos teóricos versados nos estudos sobre a IA, e descrever as principais funcionalidades de sistemas de IA. Freitas (2016) partiu do pressuposto de que as pesquisas sobre a IA ainda não se desenvolveram no meio acadêmico-científico brasileiro. Pesquisas junto aos mais diversos catálogos de artigos, teses, dissertações e livros, tanto em meio impresso quanto eletrônico, revelam a completa ausência de trabalhos científicos sobre a IA em língua portuguesa do Brasil. Apesar disso, a temática tem sido acolhida por diversas áreas de pesquisa no exterior, tais como Ciências da Computação, Linguística, Processamento de Fala e Inteligência Artificial (PÖCHHACKER, 2004; LEE, 2015), cujas reflexões teóricas também apoiarão este novo trabalho.

No montante de mais de 285 artigos relacionados ao tema, os resultados de Freitas (*op. cit.*) revelam que, nos últimos 15 anos, a IA vem sendo apresentada em conferências internacionais com maior frequência, passando de um assunto estranho ao público em geral para um dos principais interesses de pesquisa para os estudiosos da área de Processamento de Fala. Segundo Waibel e Fügen (2008), a convergência de novas tecnologias e a crescente atenção à necessidade de se ter uma comunicação

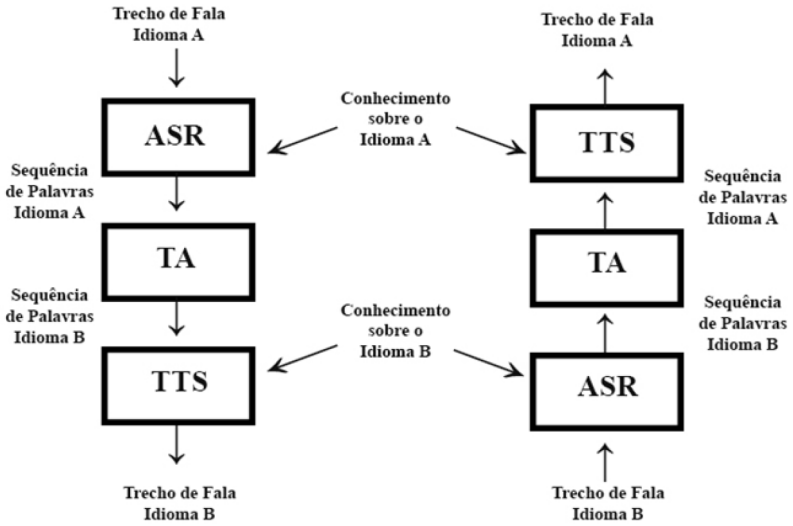
multilíngue em um mundo globalizado são os motivos que explicam o crescente interesse em tecnologias desse tipo.

Os precursores na história da implementação da IA, todavia, são das Ciências da Computação, cujos esforços remontam às primeiras atividades humanas relacionadas à matemática e mecanização do pensamento humano (FONSECA FILHO, 2007). A partir dos primeiros computadores, até o surgimento de *apps* de tradução de fala, a ideia de usar computadores para derrubar as barreiras linguísticas que separam os povos tem sido um dos principais focos das pesquisas das Ciências da Computação (FREITAS, 2016).

O levantamento bibliográfico de Freitas (*op. cit.*) concernente aos estudos sobre IA revelou que os teóricos da área, em cenário internacional, utilizam os seguintes termos para definir conceitualmente esse tipo de tecnologia: *speech translation*, *speech-to-speech translation, spoken language translation* e *simultaneous translation.* Em relação aos termos adotados em língua portuguesa, verificou-se que os termos "tradução de voz" é utilizado pelas pesquisas brasileiras, embora ainda evidentemente escassas, e que "tradução de fala" é utilizado por estudiosos de Portugal, em correspondência ao termo *speech translation* (BARREIRO *et al.*, 2014; DUARTE, 2014; PINTO, 2015).

Caracterizado como um sistema que realiza a tradução de trechos de fala de uma língua para outra, a pesquisa de Freitas (*op. cit.*) constatou que a IA só é possível graças a três componentes tecnológicos que são integrados em uma só interface: reconhecimento automático de fala (ASR), tradução automática (TA) e síntese de voz (TTS). A figura 1 ilustra a arquitetura básica de um sistema de IA bidirecional para o par de idiomas A e B.

Figura 1 – Arquitetura de um sistema de IA bidirecional adaptado de Lee (2015) por Freitas (2016).



O trecho de fala do idioma A é inserido no mecanismo e é reconhecido pelo componente de ASR, a fim de se produzir um texto contendo a transcrição do trecho, ainda no idioma A. A transcrição então é traduzida pelo componente de TA que produz uma tradução textual no idioma B. Por fim, o texto traduzido passa pelo componente de TTS e então o trecho de fala contendo a mensagem traduzida é gerado e verbalizado (LEE, 2015).

Há sistemas de IA que são construídos a partir de componentes desenvolvidos de forma independente, acoplados em uma só interface através de softwares específicos (AIKEN; SIMMONS; BALAN, 2010). Na maioria das vezes, todavia, este tipo de acoplagem não produz feedback sobre os procedimentos isolados. Consequentemente, torna-se impossível que um erro seja corrigido pelos componentes antes do erro gerado pelo ASR ser replicado para os demais componentes de forma sequencial, comprometendo o desempenho de todo o sistema (LEE, 2015). Por outro lado, há sistemas de IA desenvolvidos a partir de uma integração mais sofisticada entre os componentes, visando um processamento coerente e um desempenho otimizado de ponta a ponta. Em sistemas deste tipo, o componente de ASR produz várias opções (hipóteses) para que o componente de TA

possa escolher a que melhor se encaixa ao contexto do discurso com que se trabalha (CASACUBERTA *et al.*, 2008).

Hashimoto *et al.* (2012) afirmam, no entanto, que poucos sistemas têm proposto este tipo de abordagem para o componente de TTS, o que pode prejudicar a compreensão do conteúdo traduzido e verbalizado pelo sistema. Segundo Barreiro *et al.* (2014), o estado da arte atual da IA caracteriza-se por:

> uma integração relativamente fraca entre os três módulos, não explorando as sinergias existentes entre o reconhecimento e a tradução, entre a tradução e a síntese e ainda entre o reconhecimento e a síntese. Por exemplo, o módulo de reconhecimento escolhe normalmente uma única hipótese de transcrição que será a entrada do módulo de tradução. Se, em alternativa a esta hipótese, for oferecida ao módulo de tradução uma lista de possíveis transcrições, este pode decidir qual a mais adequada aos modelos de tradução. Por outro lado, o módulo de síntese assume que receberá como entrada um texto fluente, o que usualmente não acontece quando essa entrada resulta de um módulo de tradução automática (BARREIRO *et al.*, *op. cit.*, p. 78).

No intuito de comprovar as principais funcionalidades desse sistema, Freitas (2016) implementa também um experimento com a ferramenta de tradução automática Google Tradutor, lançada em 2006. Além de ser gratuita, a ferramenta é amplamente utilizada na internet por usuários comuns e profissionais das mais variadas áreas. Segundo informações da empresa subsidiária Google, a ferramenta que suportava apenas os idiomas inglês, espanhol, francês e alemão, no período de 2006 a 2008, no começo de 2009 já atingia o número de 41 idiomas, ou seja, 98% dos idiomas lidos na internet.

Construído a partir de abordagens estatísticas de TA, como a do sistema em código aberto Moses, muito usado pela comunidade de pesquisadores e sistemas comerciais, a grande vantagem do Google Tradutor é o acesso a uma grande quantidade de *corpora* paralelos disponíveis na internet, o que torna possível a tradução de um número elevado de pares de línguas. De acordo com Barreiro *et al.*

(2014), a qualidade dessas traduções, todavia, depende de fatores como a proximidade entre a língua-fonte e a língua-alvo, ou a quantidade e qualidade dos *corpora* disponíveis para a tradução dos pares de línguas. Atualmente, além da tradução automática de fala utilizada nos experimentos do trabalho de Freitas (2016), a ferramenta oferece serviços de tradução de textos, imagens, websites e documentos em vários formatos para mais de 100 idiomas. Disponível no formato de *app* para *smartphones, tablets* e *desktops*, funciona tanto online quanto offline, possibilitando que os usuários baixem determinados pares de idiomas a serem armazenados na memória dos dispositivos.

A função de tradução automática de fala do Google Tradutor no navegador Google Chrome faz uso do recurso Web Speech API, tecnologia ASR desenvolvida na linguagem Java e integrada ao navegador no começo de 2013. Essa tecnologia é utilizada também por desenvolvedores para integrar a ASR aos seus aplicativos online (HASHIMOTO *et al.*, 2012).

À ferramenta, Freitas (*op. cit.*) submeteu o discurso de posse do 44º presidente dos EUA, Barack H. Obama, proferido na manhã do dia 21 de janeiro de 2013, na cerimônia de posse presidencial, realizada tradicionalmente em frente ao Capitólio dos Estados Unidos, prédio que sedia o poder legislativo do governo norte-americano, em Washington, nos Estados Unidos. A ferramenta também foi testada acoplando-se a ela um diálogo entre uma recepcionista de um hotel e um hóspede, que se caracteriza pela presença de nomes próprios, números cardinais e ordinais, perguntas, respostas, horários etc. Para os estudiosos da IA, um discurso político e outro voltado ao turismo, por conterem características próprias, podem representar desafios distintos à tecnologia da IA (FÜGEN, 2008; GRAZINA, 2010).

Segundo Freitas (2016), o experimento revelou que, embora o Google Tradutor não seja uma ferramenta de tradução de fala especializada (discurso político ou voltado ao turismo), a tecnologia ASR se saiu melhor com o reconhecimento de fala do diálogo entre a recepcionista de hotel e o hóspede. Quanto ao discurso de posse de

Barack Obama, o sistema transcreveu 50% dos trechos de forma adequada e 50% de forma inadequada, o que sugere que por mais que a ferramenta possa ser usada para a tradução de diversos tipos de textos orais, um dos fatores mais relevantes é a qualidade do áudio inserido na tecnologia. Nos casos estudados por Freitas (2016), o discurso de Obama continha muitos ruídos, como gritos da plateia, salva de palmas etc., ao passo que o discurso entre um hóspede e uma recepcionista de hotel encontrava-se menos ruidoso.

Buscando igualmente investigar os estudos concernentes à IA, o trabalho de Freitas e Esqueda (2017) apresenta um levantamento bibliográfico sobre esta tecnologia, com vistas a construir um inventário de candidatos a termos (FINATTO, 2002; FROMM, 2005) que a denominam e conceituam, para seu possível uso em português do Brasil.

Os autores tomaram como base o mesmo *corpus* levantado por Freitas (*op. cit.*) para a construção de um vocabulário terminológico que representasse os termos mais frequentemente utilizados na literatura consultada, almejando revelar como os autores de fato denominam a IA e as tecnologias utilizadas para operá-la. A partir de um *corpus* de 285 artigos científicos (Freitas, *op. cit.*) foi possível elaborar um inventário de candidatos a termos em que Freitas e Esqueda (*op. cit.*) verificam que *speech translation*, *speech-to-speech translation*, *spoken language translation* e *simultaneous translation* são termos, em inglês, recorrentes na literatura que define e discute essa tecnologia. Em relação aos termos adotados em língua portuguesa, os autores constataram que o termo "tradução de voz" é utilizado pelas pesquisas brasileiras, embora ainda evidentemente escassas, e que "tradução de fala" é utilizado por estudiosos de Portugal, em correspondência ao termo *speech translation*.

Assim, esses trabalhos anteriores, de Freitas (2016) e Freitas e Esqueda (2017), buscaram dar os primeiros passos rumo a um estado da arte sobre a IA, descrição de seus sistemas e vocabulário mais comumente utilizado para descrevê-los. Almejando dar continuidade a essa investigação, o estudo que ora se apresenta tem como

proposta investigar como os sistemas de IA, que revelam inconsistências com interpretações muito literais e errôneas, têm buscado capturar a intenção do discurso, vislumbrando interpretações mais consistentes.

3. A intenção discursiva

Segundo Haugh e Jaszczolt (2012), o estudo da intenção dos locutores tomou parte no pragmatismo contemporâneo por meio de três abordagens que, embora sejam distintas, se inter-relacionam de algum modo. A primeira delas remonta à filosofia medieval e às investigações sobre a lógica contextual, levando ao estudo da intenção. A segunda tem início na década de 1950, com a filosofia da linguagem e as tentativas de definir o significado através do uso da linguagem, que culminou no emprego do conceito de "efeito pretendido" do ato comunicativo. A terceira abordagem, que segundo Haugh e Jaszczolt (*op. cit.*) tornou-se a mais influente, caracteriza-se pela tentativa de retomar análises semânticas formais a partir do emprego dos conceitos de significado, "mensagens pretendidas" e "conteúdo comunicado".

Embora não seja o objetivo deste estudo tratar da intenção no âmbito dos estudos linguísticos ou da filosofia da linguagem, vale ressaltar que a investigação sobre os elementos intencionais é o principal foco da prosódia e da pragmática, áreas que ocupam diferentes correntes de estudo no interior da área da Linguística e que reúnem conceitos relativos à psicologia da fala, sociologia da fala, fonologia e demais áreas correlatas. Pöchhacker (2004) afirma que, ao lado da entonação, componentes prosódicos são bastante relevantes para a percepção e compreensão do processo de interpretação.

A partir do entendimento do conceito de intenção como um conjunto de estados mentais, é possível afirmar que todas as línguas, na qualidade de veículos desses estados (HAUGH; JASZCZOLT, 2012), cada qual com uma intensidade diferente, permitem ao falante empregar em seu discurso aquilo que deseja

comunicar de forma mais ou menos consciente. Em línguas como o japonês, por exemplo, o estilo do discurso oral difere bastante do estilo empregado na escrita. A expressão oral japonesa é fragmentária e nela inclui-se, direta ou indiretamente, a intenção do locutor (MORIMOTO *et al.*,1992).

Assim, a fala humana produzida espontaneamente é repleta de hesitações, repetições, pausas e expressões agramaticais, que lhe conferem um caráter "degenerado" (CHOMSKY, 1976), contendo ainda intenções que visam atender determinados objetivos do falante. E para que os sistemas de IA sejam eficientes e realizem a tarefa a que se propõem, é necessário que sejam capazes de lidar exatamente com esse tipo de conteúdo (BARREIRO *et al.*, 2014).

O processamento de fala humana por sistemas computacionais é particularmente difícil devido à existência de elementos linguísticos e extralinguísticos que não estão diretamente disponíveis nas transcrições de áudio utilizadas para a elaboração dos sistemas. E esse problema não está restrito apenas aos computadores, pois, assim como apontado por Labov (2008), representa um desafio até mesmo para os linguistas.

No universo de processos inconscientes que constituem a fala humana, concentramos nossa análise nos aspectos concernentes à intenção do falante e ao modo com que ela vem sendo abordada pelos sistemas de IA, foco deste trabalho.

Ao apresentar um sistema de IA que faz uso da intenção do locutor, Morimoto e Kurematsu (1993) declaram, como veremos a seguir, que a prosódia desempenha um papel importante na transmissão de informações extralinguísticas, tais como a intenção de um falante. A partir de fenômenos de linguagem como esse, aparecem enunciados fragmentários, fortemente contextuais, inversões, repetições, redistribuições, expressões agramaticais etc. Além disso, os autores ressaltam que a linguagem oral, nesse caso, a língua japonesa, comumente omite elementos que podem ser facilmente inferidos a partir do contexto, como no caso dos pronomes "eu" e "você" (em japonês).

Iida, Sumita e Feruse (1996) também estudam como os sistemas de TA e, como consequência, de IA, podem conceder à tradução da linguagem falada um tratamento mais adequado. Para os autores, tais sistemas devem extrair o significado de uma sentença, com valores padronizados que possam executar as inferências heurísticas, que por sua vez são bastante eficazes em explicar a intenção e o conteúdo proposicional da fala, por meio de uma palavra ou frase-chave.

Hong, Koo e Yang (1996) tratam das elipses do discurso oral marcadas por omissões de palavras ou sentenças, permanecendo seu significado subentendido (como no exemplo em "saímos ontem à noite", com elipse de "nós") e dificultando o processamento de fala por sistemas computacionais.

Para Yang e Park (1997), um sistema pode ser criado apenas para fornecer ao usuário a intenção parcial da fala, sendo que caberá a ele inferir o restante da mensagem, pressuposto esse também valorizado no estudo de Blanchon e Boitet (2000).

Buscando descrever uma nova abordagem que combina a análise baseada na gramática e em nível frasal, o trabalho de Langley (2002) estuda um sistema que transforma os enunciados em uma representação de interlíngua semântica de uma língua a outra, que seria parcialmente suficiente para que o falante descobrisse a intenção de seu interlocutor.

Na perspectiva dos estudos sobre a intenção, Zong e Seligman (2006) argumentam que quanto mais as falas traduzidas pelos sistemas forem abrangentes, isto é, fora de um domínio específico, mais os usuários deverão cooperar e se ajustar aos resultados exibidos pelos programas, para entender a intenção da fala em questão, refletindo sobre o que ouvem, resolvendo e esclarecendo determinadas ambiguidades.

Devido à importância da intenção para a comunicação oral e o interesse dado ao tema pelos pesquisadores da IA, este estudo tem por objetivo realizar uma revisão bibliográfica em busca da descrição das características gerais de sistemas

que buscam capturar a intenção do falante. Os sistemas que serão aqui descritos foram desenvolvidos com o intuito de, ao traduzir a linguagem oral espontânea entre falantes de diferentes línguas, capturar as variações acústicas expressas por eles, suas intenções e emoções.

4. Materiais e métodos

Este estudo constitui-se de uma revisão da literatura especializada, realizada a partir do *corpus* já levantado sobre o tema em Freitas (2016) e Freitas e Esqueda (2017). Com vistas à atualização dos dados, procedeu-se a um novo exame junto aos bancos de dados utilizados para a compilação do *corpus*. Novas pesquisas no Google Acadêmico geraram o total de outros 36 artigos, que somados aos artigos que já compunham o *corpus* (285), totalizam 321 artigos, conforme mostra a tabela 1.

Tabela 1 – Composição atualizada do *corpus* de pesquisa sobre IA até 2018.

Período	Base de dados	Quantidade
1987-2018	Google Acadêmico	106
	IEEE Xplore	141
	interACT	53
	ACM-DL	17
	E-mail	3
	Correio terrestre	1
	Total	321

Após a atualização do *corpus*, realizou-se a busca por artigos que tratam da captura da intenção discursiva dos locutores nos sistemas de IA. As palavras-chave utilizadas, em inglês, para a busca foram *speaker's intention*, *intention* e *intended message*.

Os critérios de inclusão para os estudos encontrados foram a aplicação de

abordagens para a captura da intenção discursiva, a descrição dessas abordagens e a presença de discussões sobre o uso da intenção discursiva para o aprimoramento de sistemas de IA. Foram excluídos estudos que relatam a utilização dessas abordagens nas tecnologias de ASR, TA e TTS separadamente. Essas tecnologias tomadas de forma isolada não configuram um sistema de IA (FREITAS; ESQUEDA, 2017).

Em seguida, buscou-se estudar e compreender as características gerais desses sistemas, suas arquiteturas e mecânica de funcionamento, para que então se pudesse proceder à composição da revisão bibliográfica proposta para o presente estudo.

5. Como os sistemas de IA capturam a intenção do falante?

Do montante de 321 artigos, foram encontrados 10 artigos que versam sobre a captura da intenção discursiva do falante e apenas um artigo que faz menção à referida expressão, sem todavia tratar da sua aplicação em sistemas de IA. Os artigos encontrados são compostos por relatos de experimentos, apresentações de protótipos e descrições de abordagens aplicadas à IA.

Zong e Seligman (2006) afirmam que na busca por interpretações automáticas mais realistas, que se comparem a interpretações humanas ou até mesmo as superem, é preciso que esses sistemas sejam capazes não somente de produzir interpretações corretas, mas, dentre outras questões, esclarecer a intenção discursiva dos usuários de forma interativa e cooperativa. Segundo os autores, em busca dessa interatividade, várias abordagens foram propostas desde os anos de 1990.

As características gerais, as arquiteturas e mecânicas e a evolução dos sistemas descritas pelos estudos utilizados para a composição desta revisão bibliográfica são apresentadas na seção a seguir em ordem cronológica, que parte do ano de 1992 até o presente.

5.1 Características gerais

O estudo de Morimoto *et al.* (1992) apresenta uma visão geral do sistema SL-

TRANS2, um protótipo desenvolvido para interpretar trechos de fala do japonês para o inglês. Trata-se de um sistema de interpretação telefônica cujo domínio discursivo estava limitado a um único diálogo relacionado ao tema "inscrição de participantes em conferências internacionais".

Embora um tanto limitado no que diz respeito à integração entre as tecnologias de reconhecimento de fala, tradução automática e síntese de voz, o sistema SL-TRANS2 possui algumas características que o distinguem dos demais sistemas da época. Segundo os desenvolvedores, os trechos de fala contínua em japonês são reconhecidos com alto índice de precisão, mesmo ao utilizar uma técnica de ASR independente de locutor, que embora tenha o objetivo de ser mais abrangente, pode apresentar inúmeras inconsistências.

A característica de maior relevância para o presente estudo, todavia, é o fato de o SL-TRANS2 ter sido desenvolvido com a capacidade de capturar a intenção dos usuários e utilizá-la para os processos de tradução automática e síntese de voz.

Em outro estudo desenvolvido no mesmo laboratório, Morimoto e Kurematsu (1993) descrevem o ASURA[4], um sistema de interpretação telefônica baseado no SL-TRANS2, que além de também ser independente de locutor, interpreta trechos de fala do japonês para o inglês.

Diferentemente dos dois estudos citados acima, Hong, Koo e Yang (1996) apresentam a descrição de um analisador morfológico desenvolvido para ser empregado no processamento de linguagem para sistemas de IA. Enquanto os demais autores apresentam sistemas, este estudo discorre sobre um tipo de abordagem através da qual os sistemas de IA podem ser treinados.

O analisador morfológico foi desenvolvido para lidar com trechos de fala coreana a partir da modificação de um algoritmo denominado CYK (Cocke-Younger-

[4] À versão final do ASURA, todavia, foi implementado um sistema de TTS diferente daquele citado pelos autores.

Kasami), que é capaz de determinar se uma cadeia de caracteres pode ser gerada por uma determinada gramática livre de contexto e, caso seja possível, como ela pode ser gerada. Segundo os autores, a utilização desse algoritmo torna possível a análise da grande quantidade de fenômenos que ocorrem na fala espontânea, como elipses, palavras curtas, mal pronunciadas e pouco audíveis etc. Nesse prisma, para lidar com essas características do discurso oral, é necessário estudar a língua e a cultura, como é o caso do coreano.

Como essa abordagem é baseada em regras, foi necessário construir um conjunto de 112 regras de conexão e sete dicionários, compostos por cerca de 81.000 palavras-chave. Segundo os autores, a abordagem alcança uma taxa de sucesso de 93,0%. O *corpus* utilizado nos testes foi compilado a partir de trechos de diálogos relacionados ao domínio do turismo, mais precisamente na situação comunicativa de determinado cliente realizar a reserva de um quarto de hotel.

Assim como o estudo citado anteriormente, o trabalho de Iida, Sumita e Furuse (1996) não apresenta o desenvolvimento de nenhum sistema de IA em particular. O que os autores trazem no estudo é uma descrição da abordagem de tradução de fala através do uso de exemplos. Segundo os autores, existem sete requisitos para o sucesso da tradução de fala: i. processamento incremental; ii. manipulação do discurso oral; iii. manipulação de expressões eufemísticas; iv. processamento determinístico; v. velocidade suficiente para evitar interrupções durante o ato comunicativo; vi. tradução de alta qualidade e vii. correção de erros de reconhecimento de fala.

Após enumerar os sete requisitos, os autores expõem quais são as abordagens mais utilizadas para atendê-los e os métodos que estão se tornando tendência na área. Ao relacionar o estudo a sistemas já implementados Iida, Sumita e Furuse (1996) citam o sistema Galaxy, que possibilita a interpretação automática online de serviços de informação para viajantes. O sistema faz uso da representação de *frames* semânticos de modo a transformar um trecho de fala em uma expressão concreta e simples que esteja em conformidade com uma das representações internas do sistema e facilite a

manipulação do significado do trecho de fala. Essa abordagem é conhecida como interlíngua e gera uma espécie de paráfrase do trecho de fala.

Segundo os autores, a captura do significado de uma frase é feita por meio de inferências heurísticas, ou seja, decisões não racionais que requerem uma grande quantidade de cálculos. Nesse sentido, a inferência é muito eficiente para explicar a intenção discursiva de um falante e o conteúdo proposicional da frase através de frases ou palavras-chave.

A conclusão a que Iida, Sumita e Furuse (*op. cit.*) chegam é que essa abordagem, por demandar maior desempenho tecnológico, pode funcionar bem em sistemas especialistas. Em sistemas mais abrangentes, no entanto, a eficiência diminui.

A abordagem interlíngua também foi empregada no desenvolvimento de um protótipo de IA do Instituto de Pesquisas em Eletrônica e Telecomunicações - ETRI, na Coreia do Sul (YANG; PARK, 1997). O protótipo interpreta do coreano para inglês e para japonês. O sistema traduz trechos de fala que contenham discurso no domínio de planejamento de viagem com vocabulário de 5.000 palavras.

Para esse protótipo, foi desenvolvida uma abordagem através da qual a intenção discursiva é transferida de um usuário para outro, de modo que os usuários possam compreender a interpretação mesmo quando o sistema não for capaz de realizar a tradução de fala corretamente. Segundo Yang e Park (*op. cit.*), isso se dá pela possibilidade de os usuários inferirem as mensagens através tanto do contexto quanto de suas próprias inteligências.

Esta abordagem parece bastante ambiciosa no sentido de que, embora assuma que a tradução de fala possa conter erros,

> o objetivo é alcançar um alto desempenho de ponta a ponta, ou seja, um desempenho que se equipare ao desempenho de intérpretes humanos, em contraste com sistemas convencionais, que buscam

> apenas altos desempenhos nos extremos do processo (YANG; PARK, *op. cit.*, p. 87, tradução nossa)[5].

Outro ponto distintivo é que o protótipo também oferece suporte a outros canais multimídia, que podem ser utilizados para produzir resultados mais bem-sucedidos do que os observados através do uso de mídia traduzida a partir de trechos de fala.

Devido ao fato de os desenvolvedores estarem preocupados com um alto desempenho de ponta a ponta e fazerem uso da transferência de intenção, eles desenvolveram uma metodologia de avaliação que se propõe a mensurar a compreensão da intenção do falante, classificando a qualidade da interpretação em três níveis:

> A: o usuário entende perfeitamente a intenção do falante;
> B: o usuário entende a intenção do falante apesar dos pequenos erros; e
> C: o usuário não consegue entender.

Mensurar a intenção discursiva, todavia, é uma tarefa que requer o emprego de uma série de variáveis subjetivas, o que torna a mensuração muito mais dispendiosa em termos tecnológicos do que as metodologias que fazem uso de cálculos probabilísticos. Por isso, os desenvolvedores do protótipo também aplicaram uma metodologia de avaliação mais objetiva, tendo por base a teoria da informação de Claude Shannon (1948).

Blanchon e Boitet (2000) apresentam um sistema desenvolvido pelo grupo CLIPS++, membro do consórcio C-STAR II. O sistema foi desenvolvido através da integração de sistemas de ASR, análise linguística, geração linguística e TTS, voltados

[5] *The goal of our approach is to achieve a high end-to-end, i.e., human-to-human performance in contrast to those of most conventional speech translation systems pursuing only high input-to-output performances* (YANG; PARK, *op. cit.*, p.87).

para tarefas específicas. Utilizado em demonstrações, sem o objetivo de ser comercializado ou reproduzido em larga escala, o sistema possui uma abordagem que possibilita a cooperação entre os sistemas que o compõem e a convergência de tecnologias desenvolvidas por outros grupos do consórcio.

As tecnologias desenvolvidas por eles foram incrementadas para operarem apenas com o francês. Portanto, para realizar traduções para outras línguas, é preciso que esses sistemas sejam treinados e adaptados. Para desenvolver as tecnologias, os pesquisadores empregam uma abordagem denominada "formato de interface" (IF), que depende de atos de fala, conceitos e argumentos (LEVIN *et al*., 1998).

Tendo por base a explicação de Blanchon e Boitet (2000) do IF, enquanto os atos de fala descrevem a intenção, o objetivo e a necessidade do falante, os conceitos definem o foco do ato de fala. Vários conceitos podem aparecer em um IF. Argumentos são valores das variáveis discursivas. Os autores exemplificam por meio da frase "na semana do dia 12 temos quartos individuais e duplos disponíveis", pronunciada por um agente hoteleiro, o seguinte IF deve ser produzido: a:give-information+availability+room (roomtype=(single; double), time=(week, md12)).

Outra etapa é o processamento de contexto, em que, segundo Blanchon e Boitet (2000), três pontos são focalizados: o contexto global, o contexto dialógico e o contexto linguístico:

> O contexto global contém, pelo menos, o tipo de diálogo, as características dos participantes, em específico seus nomes, sexo, idades e nível relativo de cortesia, suas intenções, quando disponíveis, e talvez os nomes de seus locais, já que podem ser personificados (BLANCHON; BOITET, *op. cit*., p.4, tradução nossa)[6].

[6] *The global context contains at least the type of dialogue, the characteristics of the participants, in particular their names, sex, ages and relative politeness level, their intentions if available, and perhaps the names of their locations, because they can be personified* (BLANCHON; BOITET, 2000, p.4).

A importância dessa cadeia de elementos contextuais reside no fato de que não somente os sistemas precisam lançar mão deles, mas, assim como Blanchon e Boitet (*op. cit.*) afirmam, os intérpretes humanos também precisam desse tipo de informação.

O sistema também realiza processamento prosódico, cujo objetivo é a obtenção de marcas prosódicas a serem passadas para os analisadores linguísticos. Essas marcas prosódicas também podem ser usadas pelos geradores em conjunto com outras características semânticas e pragmáticas, tais como a intenção do falante, para produzir resultados mais adequados à situação de fala e que contenham *tags* usadas pelos sintetizadores de voz para gerar uma prosódia adequada.

A utilização de *tags* também faz parte da abordagem de Langley (2002), que captura a intenção discursiva por meio da análise dos atos de fala e combina análise baseada em gramática em nível frasal e classificação automática para a IA. Além disso, essa abordagem tem características em comum com as demais abordagens citadas neste estudo, tanto por tratar-se da apresentação de um analisador linguístico quanto por estar voltada para tarefas específicas, fazer uso de representação interlíngua e lançar mão do formato IF.

Desenvolvida para trabalhar com o inglês e o alemão, a abordagem híbrida de Langley (2000) foi empregada em vários sistemas multilíngues, incluindo o sistema NESPOLE!, voltado para a interpretação automática de trechos de fala relacionados a comércio eletrônico, viagens e turismo. A abordagem é inovadora no sentido de que apresenta características híbridas, como o próprio autor afirma, e foi desenvolvida para "propiciar análises precisas em tempo real e melhorar a robustez e a portabilidade para novos domínios e idiomas" (LANGLEY, *op. cit.*, tradução nossa, p.1)[7].

[7] *The goal of this hybrid approach is to provide accurate real-time analyses and to improve robustness and portability to new domains and languages* (LANGLEY, *op. cit.*, p. 1).

Um outro estudo apresenta um sistema voltado para a interpretação automática em tempo real instalado em plataformas portáteis. Trata-se do estudo de Waibel *et al.* (2003), que descreve o Speechlator, um sistema bidirecional capaz de interpretar trechos de fala do inglês para o árabe e do árabe para o inglês, voltado para o domínio entrevistas médicas.

Também baseada em representação interlíngua, a interpretação realizada pelo Speechlator fundamenta-se na intenção discursiva em contraposição ao significado literal. A intenção discursiva é representada como um ato de fala independente de domínio, seguido de conceitos dependentes do domínio, combinados em uma representação denominada pelos autores "ação de domínio"[8].

Já o estudo de Zong e Seligman (2006) não apresenta nenhum sistema ou abordagem específicos e sim um apanhado dos principais sistemas de IA desenvolvidos até aquele momento. O estudo fundamenta-se no pressuposto de que o momento certo para o desenvolvimento de sistemas práticos de IA havia chegado. Segundo os autores, o desenvolvimento de tais sistemas nos próximos anos dependia do entendimento de que, mediante o estado da arte da tecnologia de IA, "os usuários devem cooperar e comprometer-se com os programas" (ZONG; SELIGMAN, 2006, p.114, tradução nossa)[9].

Partindo deste princípio, Zong e Seligman (2006) afirmam que os sistemas de IA podem ser classificados em uma escala que tenha por medida o grau de cooperação ou compromisso que exigem dos usuários:

> De modo geral, quanto mais ampla for a cobertura temática ou linguística pretendidas por um sistema, maior a demanda de

[8] *Domain action* (WAIBEL *et al.*, 2003).

[9] (...) *users must cooperate and compromise with the programs* (ZONG; SELIGMAN, *op. cit.*, p. 114).

> cooperação ou compromisso do usuário (ZONG; SELIGMAN, *op. cit*., p.114, tradução nossa)[10].

O princípio da cooperação e compromisso dos usuários é conhecido pela expressão "engenharia de fatores humanos"[11] e difere da abordagem empregada nas ferramentas de tradução assistida por computador (CAT tools - *Computer Assisted Translation Tools*), pois enquanto estas referem-se à tradução humana assistida por computador (MAHT - *Machine-Aided Human Translation*), aquela é a tradução automática assistida por humanos (HAMT - *Human-Aided Machine Translation*).

Na perspectiva de Zong e Seligman (*op. cit.*), o esclarecimento da intenção discursiva dos locutores, junto de elementos como ambiguidades lexicais ou estruturais, é fundamental para o desenvolvimento de sistemas de interpretação automática realistas, que almejam superar a eficiência dos intérpretes humanos. Nesse sentido, os autores argumentam que a obtenção de traduções corretas não é suficiente para a realização deste projeto, haja vista que os intérpretes humanos, além de produzirem traduções corretas, são capazes de produzir traduções interativas,

> ou seja, quando não podem traduzir diretamente um trecho de fala devido à presença de expressões incompletas ou outros problemas, eles (*os intérpretes*) normalmente pedem que o palestrante repita ou forneça esclarecimentos posteriores (1996 a,b, BOITET *apud* ZONG; SELIGMAN, 2006, p.123, tradução nossa, grifo nosso)[12].

O estudo de Cho, Ha e Waibel (2013), por sua vez, menciona a expressão *speaker's intention* uma única vez ao referir-se à tradução de uma frase alemã. Trata-se de um estudo sobre a detecção de disfluências na fala através de marcadores de

---

[10] *In general, the broader the intended linguistic or topical coverage of a system, the more user cooperation or compromise it will presently require* (ZONG; SELIGMAN, *op. cit*., p. 114).

[11] *Human factors engineering* (ZONG; SELIGMAN, *op. cit*., p. 114).

[12] *That is, when unable to translate directly an utterance due to ill-formed expressions or other problems, they often ask the speaker for repetition or further explanation* (1996 a, b, BOITET *apud* ZONG; SELIGMAN, *op. cit*., p.123).

discurso. O estudo desses autores contrasta com os demais artigos encontrados nas bases de dados consultadas, no sentido de que não traz qualquer esclarecimento sobre a utilização da intenção discursiva em sistemas de IA.

O estudo mais recente a compor a presente revisão bibliográfica é o apresentado por Kim e Kim (2018). Trata-se de uma abordagem desenvolvida a partir de um modelo de rede neural denominado IIIM (Modelo integrado de identificação de intenção baseado em redes neurais), que identifica as intenções discursivas dos locutores com o objetivo de solucionar o principal problema do processamento de fala: a existência de elementos linguísticos e extralinguísticos.

Segundo os autores, um sistema de IA deve ser capaz de capturar as intenções discursivas dos falantes, que podem ser representadas por combinações de atos de fala, predicadores e emoções. O erro das abordagens anteriores, afirmam os autores, foi ter tentado identificar esses elementos de forma independente, haja vista a conexão estreita entre eles.

O modelo proposto por Kim e Kim (2018) identifica simultaneamente os três elementos e os acomoda em camadas ocultas concebidas para a incorporação de abstrações informativas usadas para a identificação de outros atos de fala, predicadores e emoções. Esses elementos são entidades denominadas "nós" e, quando acomodados às camadas ocultas, são parcialmente treinados por três ciclos de *backpropagation*: i. treinamento dos nós associados à identificação da ação do discurso (como "solicitar informações ou referências", "responder ou afirmar", que podem indicar a intenção do falante em qualquer domínio), ii. identificação do predicador (como "tarde", "parte", "ser", "encorajar", que geralmente se associam ao conteúdo principal), e iii. identificação da emoção ("nenhuma", "tristeza", que expressam as emoções ou atitudes do falante).

Kim e Kim (*op. cit.*) explicam que enquanto um ato de fala e um predicador representam a intenção explícita do falante, uma emoção representa uma intenção implícita, e as duas intenções se complementam sequencialmente, de modo que esse

modelo leva em consideração o desencadeamento entre um dado ato de fala e os atos de fala anteriores.

Nos experimentos realizados com a abordagem de Kim e Kim (*op. cit.*), o modelo proposto apresenta maior pontuação do que os modelos que consideram os três elementos separadamente: 6,8% maior na identificação de fala, 6,2% maior na identificação do predicador e 4,9% maior na identificação da emoção. Com base nos resultados experimentais, os autores chegam à conclusão de que a arquitetura de integração proposta e os ciclos de *backpropagation* podem ajudar a aumentar o desempenho da identificação da intenção.

5.2 Arquiteturas e mecânica

Nos itens a seguir encontram-se descritas as arquiteturas e mecânica dos sistemas apresentados nos estudos de Morimoto *et al.* (1992), Morimoto e Kurematsu (1993), Yang e Park (1997), Blanchon e Boitet (2000) e Waibel *et al.* (2003). Os estudos de Hong, Koo e Yang (1996), Iida, Sumita e Furuse (1996), Langley (2002) e Kim e Kim (2018) não são descrições de sistemas, mas sim de abordagens e, por isso, não constam nos itens que se seguem.

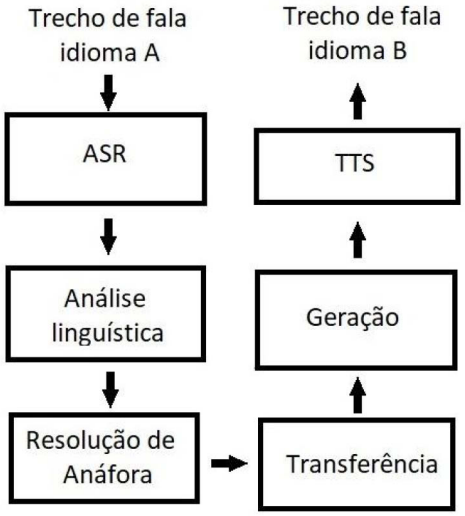
5.2.1 SL-TRANS2

O sistema SL-TRANS2, apresentado por Morimoto *et al.* (1992), realiza a interpretação automática em seis etapas, como ilustrado na figura 2. Após as etapas de ASR, análise linguística e resolução de anáfora, o sistema captura a semântica do trecho de fala japonesa e o coloca em uma estrutura denominada *feature structure* (MORIMOTO *et al.*, 1992).

As duas últimas etapas são a transferência e a geração, conforme a figura 2. Segundo a descrição feita pelos autores, essa estrutura é composta de duas partes: conteúdo intencional e conteúdo proposicional. Enquanto a primeira designa a intenção do locutor, que se expressa em conceitos que independem da língua, a

segunda se expressa em conceitos que dependem da língua, ou como Morimoto e Kurematsu (1993) colocam, trata-se de uma proposição neutra.

Figura 2 – Arquitetura do sistema SL-TRANS2, adaptado de Morimoto *et al.* (1992).



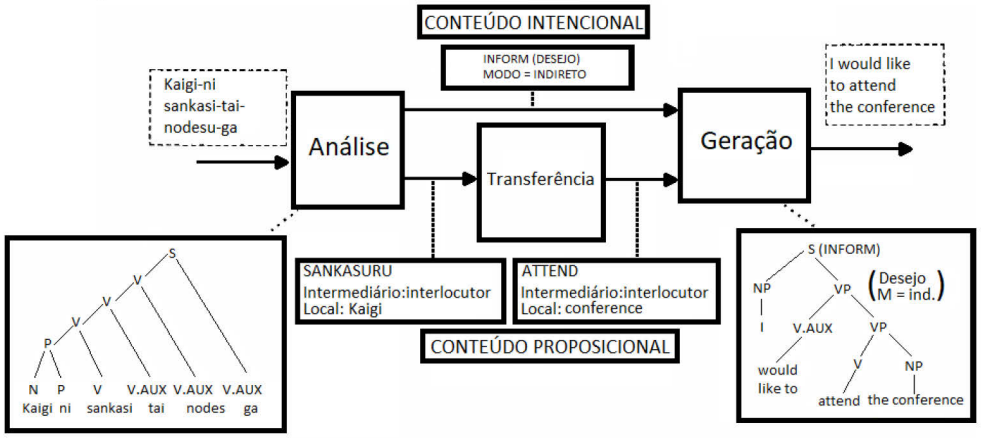
Na primeira etapa, o sistema transfere apenas o conteúdo proposicional para os conceitos da língua-alvo, nesse caso o inglês. E, na segunda etapa, os conteúdos intencional e proposicional são unidos em uma expressão final.

Tendo em vista que somente o conteúdo proposicional é transferido para os conceitos da língua-alvo, não fica claro na explicação de Morimoto *et al.* (1992) em que momento a intenção do locutor passa para a língua-alvo, nem como a utilização desse conteúdo intencional pode influenciar no resultado final da interpretação.

5.2.2 ASURA

Embora a arquitetura do sistema ASURA (MORIMOTO; KUREMATSU, 1993) seja praticamente a mesma do SL-TRANS2, nesse novo trabalho os autores passaram a se referir ao uso da intenção do locutor como "método de tradução de intenção" (ilustrado na figura 3), tendo por base o trabalho de Kurematsu *et al.* (1991). Além disso, uma explicação mais detalhada, não só da totalidade do processo de interpretação automática, bem como da captura da intenção do locutor, é oferecida através de figuras e fluxogramas.

Figura 3 – Arquitetura do sistema ASURA, adaptado de Morimoto e Kurematsu (1993).



O desenvolvimento da tecnologia de ASR requer dois tipos de modelos: modelo acústico e modelo linguístico. Para o modelo acústico do sistema de ASR do ASURA foi utilizado o Modelo de Markov Escondido (HMM – *Hidden Markov Model*). As palavras são decompostas em unidades denominadas fones, que são afetados acusticamente pelos fones presentes em seu entorno. Deste modo, centenas de modelos alofones são gerados a partir de um banco de dados de voz de grandes proporções. Para o modelo linguístico, utilizou-se uma gramática livre de contexto (GLC), cuja manutenção e adaptação mostra-se superior aos modelos linguísticos convencionais. Os dois modelos são então combinados por um analisador sintático preditivo que possibilita o processamento de fala contínua (MORIMOTO; KUREMATSU, 1993).

Para tornar o sistema independente de locutor, adotou-se uma abordagem adaptativa através do algoritmo VFS (*vector field smoothing*), em que apenas cerca de dez palavras são suficientes para adaptar o sistema à fala de um novo usuário. O sistema admite trechos de fala proferidos frase por frase, de modo que o discurso é pronunciado de forma nítida. Para lidar com tais enunciados, regras de estrutura frasal em japonês e regras gramaticais interfrasais são inseridas na tecnologia de ASR. As frases são reconhecidas como um todo e não como um aglomerado de unidades ou palavras independentes.

Essa integração entre a tecnologia de ASR e análise linguística faz com que quase todas as frases estejam sintaticamente corretas ao final do processo de reconhecimento. No entanto, Morimoto e Kurematsu (1993) reconhecem que ainda assim existem várias ambiguidades não resolvidas pela tecnologia de ASR e isso se deve ao fato de que durante o processo são usadas somente restrições sintáticas. E, para resolver o problema, a tecnologia de ASR não gera apenas a melhor hipótese, mas várias hipóteses. Na etapa seguinte, essas hipóteses são analisadas e a que melhor satisfizer as restrições sintáticas, semânticas ou mesmo pragmáticas, é escolhida.

O método de tradução de intenção só é utilizado na etapa de TA, em que os trechos de fala são processados por um analisador baseado em uma gramática sintagmática nuclear (HPSG), que unifica e formaliza a estrutura. Para cada vocábulo lexical são definidas regras sintáticas, semânticas e pragmáticas. Segundo Morimoto e Kurematsu (*op. cit.*), o maior problema dessa abordagem é a ineficiência causada pela operação de unificação. A introdução de regras de GLC ou a implementação de um algoritmo de unificação são esforços para a solução desse problema. E, em decorrência desses esforços, o tempo de processamento foi drasticamente reduzido.

A etapa de transferência é composta de três fases: resolução de anáfora, determinação do tipo de força ilocucionária e conversão da semântica do idioma A para a semântica do idioma B. Para compreender a importância da resolução de anáfora é preciso levar em consideração que, no discurso oral em japonês, palavras que não são facilmente inferidas a partir do contexto são comumente omitidas. Pronomes pessoais como "eu" e "você" raramente são pronunciados explicitamente. Segundo Morimoto e Kurematsu (*op. cit.*), muitas vezes as anáforas podem ser resolvidas pelo uso de informações pragmáticas, como as expressões honoríficas presentes na frase. A fase de determinação do tipo de força ilocucionária é feita através da análise do conteúdo intencional e a conversão da semântica do

idioma A para a semântica do idioma B é feita através do conteúdo proposicional.

O componente final da TA, denominado geração, admite estruturas semânticas que descrevam tanto o tipo de força ilocucionária quanto o conteúdo proposicional. O papel desse componente é gerar uma árvore sintática que corresponda às estruturas semânticas do idioma A. Um conjunto de subárvores contendo informações semânticas é definido no sistema, sendo também definido tanto para cada estrutura básica de frase, como para cada expressão idiomática típica do idioma B.

Durante a geração, o conjunto de subárvores, que pode incluir todas as estruturas semânticas do idioma A, é selecionado e combinado pela operação de unificação. Por fim, uma cadeia de palavras lexicais é produzida na parte inferior da árvore sintática.

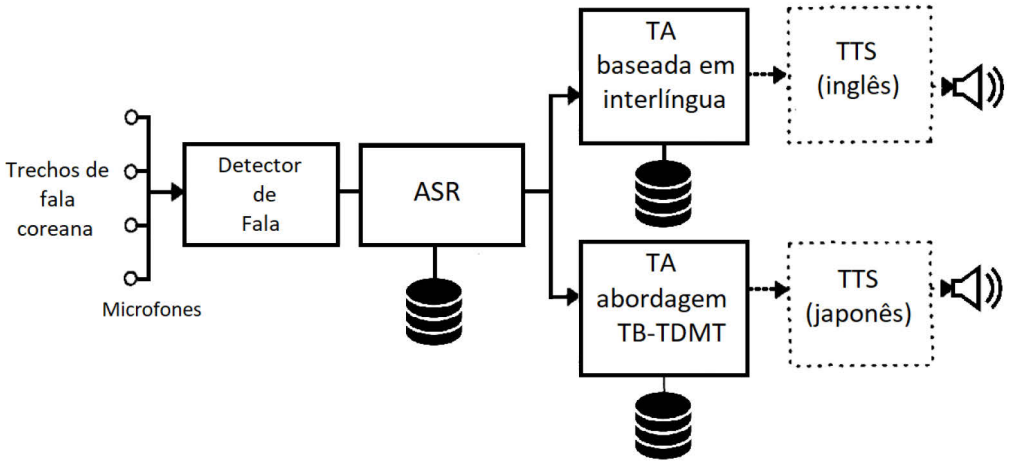
A tecnologia de TTS aplicada ao estudo de Morimoto e Kurematsu (1993) levou em consideração a necessidade de se produzir fala sintetizada nítida e natural. Segundo os autores, na TTS convencional, unidades de fala uniformes tais como CVC (consoante+vogal+consoante) ou VCV (vogal+consoante+vogal) são preparadas e o trecho de fala no idioma B é gerado pela conexão dessas unidades. O processo de conexão (concatenação) dessas unidades, no entanto, cria uma distorção e como resultado a fala sintetizada não é nem nítida, nem suficientemente natural.

Morimoto e Kurematsu (*op. cit.*) aplicam ao ASURA uma abordagem diferente, denominada Nyu-talk (1992, SAGISAKA *apud* MORIMOTO; KUREMATSU, *op. cit.*), através da qual o sistema de TTS extrai de um *corpus* de grandes proporções unidades não uniformes e as armazena em um arquivo de fala sintetizada. O sistema seleciona de forma dinâmica a combinação de unidades não uniformes menos distorcida. Na etapa final da síntese, a prosódia do trecho de fala do idioma B é controlada de acordo com a estrutura sintática da frase.

5.2.3 Protótipo de IA do ETRI

A figura 4 ilustra a arquitetura do protótipo de IA descrito por Yang e Park (1997). Para captar os trechos de fala do idioma A, o sistema utiliza um conjunto de microfones capazes de detectar fala sem uso de botões. A tecnologia de ASR transforma os trechos de fala em trechos de texto em coreano. Enquanto a tradução para o inglês lança mão da abordagem interlíngua, a tradução para o japonês lança mão da abordagem TB-TDMT (*Token-based transfer-driven machine translation*). A tecnologia de TTS para inglês e para japonês sintetiza os trechos de fala para os respectivos idiomas.

Figura 4 – Arquitetura do protótipo desenvolvido pelo ETRI, adaptado de Yang e Park (1997).



Os trechos de fala bilíngue são coletados por intérpretes humanos, que segundo os autores, atuam como sistemas de IA e controlam a dinâmica dos diálogos para que não haja sobreposição de fala. Os dados coletados são armazenados em três bancos de dados diferentes e utilizados tanto pela tecnologia de ASR quanto pelas tecnologias de TA.

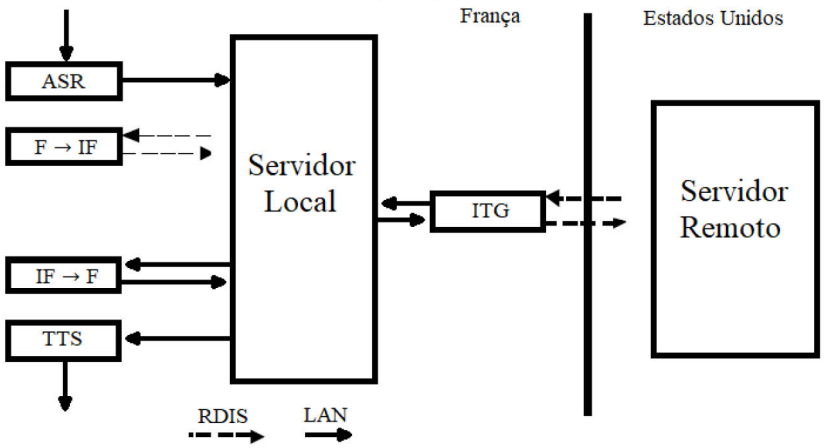
O artigo de Yang e Park (1997), todavia, não inclui detalhes sobre como os modelos acústico e linguístico da tecnologia ASR foram treinados, não há informações sobre como as abordagens interlíngua e TB-TDMT, aplicadas à TA,

desempenham a tarefa de tradução, nem de como a tecnologia TTS lida com a síntese de voz. O estudo se preocupa, em grande parte, em descrever os testes aos quais o protótipo foi submetido e quais foram os resultados obtidos através dos testes.

5.2.4 CLIPS ++

Todos os componentes do sistema desenvolvido pelo grupo CLIPS++ são servidores, que trocam basicamente dois tipos de dados: vídeo e som para videoconferência e dados que dão suporte ao processo de TA. A videoconferência é realizada por sistemas comerciais que se comunicam por meio de um protocolo de comunicação denominado telnet. Segundo Blanchon e Boitet (2000), os componentes não se comunicam entre si automaticamente, um servidor local os conecta possibilitando a comunicação (figura 5).

Figura 5 – Arquitetura do sistema desenvolvido pelo grupo CLIPS++, adaptado de Blanchon e Boitet (2000).



Os trechos de fala do idioma A são primeiramente processados pela tecnologia ASR, transformados em texto no idioma A e enviados para o servidor local através de uma conexão LAN. Do servidor local o texto é enviado para o analisador de IF através de uma rede digital com integração de serviços (RDIS). O analisador de IF transforma o texto em interlíngua. A representação interlíngua

volta para o servidor local através da mesma rede RDIS e é redirecionada através da conexão LAN até o analisador de IF que transforma a representação interlíngua em texto no idioma B, que volta mais uma vez para o servidor local e é redirecionado até a tecnologia de TTS, responsável por produzir o trecho de fala correspondente no idioma B, conforme ilustra a figura 5.

5.2.5 Speechlator

O sistema Speechlator foi desenvolvido com o objetivo de realizar interpretações automáticas em assistentes pessoais digitais (PDAs). A intenção inicial era projetar uma interface acoplada ao próprio dispositivo ou a servidores externos que poderiam ser acessados através de conexão de dados. O desempenho dos processadores instalados nos PDAs da época, todavia, não era suficiente para a implementação de tais sistemas em um curto período de tempo. Além disso, desenvolver um sistema de IA em PDAs representa um grande desafio não só em nível de desempenho, mas também devido ao fato de os microfones desses dispositivos não serem de alta qualidade. O tamanho do hardware faz com que o ruído elétrico da fonte de alimentação e da placa-mãe interfira no canal de áudio. (WAIBEL *et al*., 2003).

Como o processamento de fala também envolve o processamento de texto, a tarefa de lidar com a escrita árabe, que não inclui todas as vogais e possui sinais diacríticos apenas em determinadas tipologias textuais, é especialmente dificultosa. A escrita árabe padrão não poderia ser usada pelas tecnologias de ASR, TA e TTS. Portanto, a solução encontrada pelos desenvolvedores do Speechlator foi transcrever a escrita árabe para o alfabeto romano e, assim, criar modelos acústicos e linguísticos especiais a serem aplicados ao sistema.

O *corpus* de trechos de fala em inglês utilizado no sistema advém de um banco de dados previamente compilado e utilizado em outro sistema de IA. Os trechos foram traduzidos à mão por *experts* em árabe, que produziram cerca de 10 diferentes traduções para um mesmo trecho. Após a tradução, os mesmos *experts*

pronunciaram os trechos, que foram gravados e armazenados em um banco de dados.

A tecnologia de TA do Speechlator baseia-se em uma representação interlíngua cujo principal parâmetro é a intenção discursiva do locutor. Além disso, a interlíngua utilizada é independente de idioma, de forma a possibilitar o aporte de novos idiomas, sem afetar os idiomas já treinados. A intenção discursiva é capturada e colocada em uma representação denominada "ação de domínio" (WAIBEL *et al.*, 2003), formalizada em um documento de especificação para leitura humana e computadorizada, conforme ilustra a figura 6.

Figura 6 – Exemplo de ação de domínio, adaptado e traduzido de Waibel *et al*. (2003).

Eu tenho um esposo e dois filhos com idades de dois e onze anos.

```
dar-informação+dados-pessoais
   (família=
      spec= (conj=e,
                (cônjuge, sexo=masculino),
                (descendência,
                  quantidade=2,
                  idade=(quantidade= (conj= e, 2, 11)),
   experimentador=eu)
```

Por tratar-se de um sistema desenvolvido para PDAs, a tecnologia de TTS empregada no sistema é a Cepstral, que utiliza técnicas de pequeno porte para a seleção de unidades de subpalavra a serem sintetizadas. O usuário tem a opção de escolher entre voz masculina ou voz feminina, sem que elas estejam limitadas ao domínio. A descrição do Speechlator feita por Waibel *et al.*(2003), no entanto, não inclui ilustrações da arquitetura nem do mecanismo pelo qual o processo de IA se dá.

6. Considerações finais

A proposta deste estudo foi identificar como os autores do *corpus* investigado descrevem as formas de captura da intenção discursiva em sistemas de interpretação automática.

Além de uma retrospectiva aos trabalhos de Freitas (2016) e Freitas e Esqueda (2017), que investigam a tecnologia de IA a partir de um *corpus* composto por 285 artigos, este estudo, buscando atualizar esse *corpus*, investigou mais 36 artigos oriundos especificamente do Google Acadêmico. No montante de 321, 10 artigos tratam de forma específica da captura da intenção discursiva do falante, objeto de estudo dessa pesquisa.

Os artigos que compõem o presente estudo bibliográfico foram publicados ao longo de mais de duas décadas e meia. Nesse período, é possível observar inúmeros avanços tecnológicos que influenciaram de maneira decisiva a forma com que os sistemas de IA são projetados, integrados e empregados. No centro desses avanços encontra-se a captura da intenção discursiva dos locutores como uma das possíveis soluções para a realização do projeto de IA de qualidade.

A primeira tentativa de utilizar a intenção discursiva, descrita pelo estudo de Morimoto *et al.* (1992), surgiu cerca de uma década após a demonstração de um sistema de IA, realizada durante a convenção ITU Telecom, no ano de 1983 (NAKAMURA, 2009). Assim como os primeiros sistemas de IA, o SL-TRANS2 era um sistema de tele-interpretação, ou seja, interpretação automática através de aparelhos telefônicos. O sistema desenvolvido pelo grupo CLIPS++, por outro lado, faz uso de comunicação remota para a interpretação automática durante videoconferências (BLANCHON; BOITET, 2000).

Ao estudar a evolução dos sistemas de IA, todavia, é necessário levar em conta que cada componente (ASR, TA e TTS) pode possuir diferentes tipos de implementação. Assim, enquanto a tecnologia de TA pode ter sido desenvolvida a partir de regras (TA direta e TA por interlíngua), a tecnologia de ASR pode ser baseada em *corpus* (ASR estatística e ASR baseada em exemplos) (LEE, 2015).

Os primeiros sistemas de IA capazes de capturar a intenção discursiva foram desenvolvidos a partir da abordagem baseada em regras. O analisador morfológico apresentado por Hong, Koo e Yang (1996), no entanto, fazia uso tanto da abordagem

baseada em regras quanto da baseada em *corpus*.

A partir dos anos 2000, a abordagem mais recorrente é a baseada em *corpus*. Enquanto o sistema desenvolvido pelo grupo CLIPS++ exemplifica esse tipo de abordagem em sua forma padrão, o estudo de Kim e Kim (2018) exemplifica a utilização de *corpus* na extração de exemplos e conhecimentos a serem utilizados pelas redes neurais artificiais, um dos modelos mais frequentemente investigados nos últimos anos.

O panorama da evolução dos sistemas de IA que capturam a intenção discursiva do locutor inclui desde sistemas com nenhum tipo de feedback entre as tecnologias componentes, passando pelo surgimento da noção de que é preciso que as tecnologias colaborem entre si de forma mais ampla, promovendo a ergonomia (BLANCHON; BOITET, 2000), até as inserções teóricas sobre como o feedback e a interação entre as tecnologias e os usuários podem levar à realização de interpretações automáticas mais realistas (ZONG; SELIGMAN, 2006).

A evolução da tecnologia de IA com foco na captura do discurso, entretanto, não segue um ritmo totalmente linear. As principais abordagens ora são aplicadas com mais frequência, ora são substituídas e só retornam posteriormente junto a outras abordagens. A discussão sobre a tele-interpretação, por exemplo, presente nos primeiros estudos, volta a ser abordada somente em 2006 no trabalho de Zong e Seligman (2006).

O desenvolvimento de sistemas voltados para o processamento de fala contínua, de igual modo, está presente nos estudos de Morimoto *et al.* (1992) e Morimoto e Kurematsu (1993), na década de 1990 e, após um lapso de tempo de quase uma década, encontra-se presente no estudo de Blanchon e Boitet, nos anos 2000.

Grosso modo, para que as interpretações automáticas sejam mais realistas e se assemelhem à interpretação humana, é preciso que esses sistemas sejam capazes não somente de produzir interpretações satisfatórias, mas esclarecer a intenção

discursiva dos usuários de forma interativa e cooperativa.

Espera-se que este estudo bibliográfico possa dar continuidade a trabalhos que visem testar esses sistemas em operação, buscando fomentar as discussões acerca do que hoje se entende por sistemas de interpretação automática e formas de captura da intenção discursiva.

Referências

ACM Digital Library [Internet]. New York: ACM. 2013 - [citado em 2016 jan. 12]. Disponível em: <http://dl.acm.org/>.

AIKEN, M.; SIMMONS, L. L.; BALAN, S. Automatic Interpretation of English Speech. **Issues in Information Systems**, v. 11, n. 1, p. 129-133, 2010. Disponível em: https://pdfs.semanticscholar.org/295f/1e70392d7f16266819b603235dac7e531a5b.pdf. Acesso em: 16 abril 2018.

BARREIRO, A. *et al*. Projetos sobre tradução automática do português no laboratório de sistemas de língua falada do INESC-ID. **Linguamática**, v. 6, n. 2, p. 75-85, 2014. DOI https://doi.org/10.21814/lm.10.1.268. Disponível em: http://www.linguamatica.com/index.php/linguamatica/article/view/v6n2-6. Acesso em: 16 abril 2018.

BLANCHON, H.; BOITET, C. Speech Translation for French within the C-STAR II Consortium and Future Perspectives. *In:* SIXTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 2000, Pequim. **Proceedings**… Pequim: International Speech Communication Association, 2000. Disponível em: http://www-clips.imag.fr/geta/herve.blanchon/Pdfs/ICSLP00.pdf. Acesso em: 16 abril 2018.

BOITET, C. Dialogue-based machine translation for monolinguals and future self-explaining documents. 1996a. *In:* ZONG, C.; SELIGMAN, M. Toward practical spoken language translation. **Machine Translation**, [s.l.], v.19, n.2, p.113-137, 2005. DOI https://doi.org/10.1007/s10590-006-9000-z. Disponível em: http://www.spokentranslation.com/news/pdf/TowardPracticalSpokenLanguageTranslation.pdf. Acesso em: 16 abril 2018.

BOITET, C. Machine-aided human translation. 1996b. *In:* ZONG, C.; SELIGMAN, M. Toward practical spoken language translation. **Machine Translation**, [s.l.], v.19, n.2, p.113-137, 2005. DOI https://doi.org/10.1007/s10590-006-9000-z. Disponível em: http://www.spokentranslation.com/news/pdf/TowardPracticalSpokenLanguageTranslation.pdf. Acesso em: 16 abril 2018.

BOWEN, M; BOWEN D. Conference Interpreting: A brief history. *In:* AMERICAN TRANSLATION ASSOCIATION 25TH ANNUAL CONFERENCE, 1984, p.23, Nova York. ATA Silver Tongues. **Proceedings**... Medford, NJ: Learned Information, Inc., 1984.

CASACUBERTA, F. *et al.* Recent Efforts in Spoken Language Translation. **Signal Processing Magazine**, Maryland, v. 25, n.3, p. 80-88, 2008. DOI https://doi.org/10.1109/msp.2008.917989. Disponível em: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4490204&newsearch=true&queryText=Recent%20Efforts%20in%20Spoken%20Language%20Translation. Acesso em: 16 abril 2018.

CHO, E.; HA, T.; WAIBEL, A. Crf-based disfluency detection using semantic features for German to English spoken language translation. *In:* THE TENTH INTERNATIONAL WORKSHOP ON SPOKEN LANGUAGE TRANSLATION, Heidelberg. **Proceedings**… Heidelberg: Institute for Multilingual and Multimedia Information, 2013. DOI https://doi.org/10.1002/9781119992691.ch3. Disponível em: http://workshop2013.iwslt.org/downloads/CRFBased_Disfluency_Detection_using_Semantic_Features_for_German_ti_English_Spoken_Language_Translation.pdf. Acesso em: 16 abril 2018.

CHOMSKY, N. **Reflexões sobre a Linguagem**. Lisboa: Edições 70, 1976.

DUARTE, T. S. **Máquinas De Tradução Aplicada À Comunicação Em Tempo Real Para Desenvolvimento Distribuído De Software**. 117f. Dissertação (Mestrado) - Pontifícia Universidade Católica Do Rio Grande Do Sul, Porto Alegre, 2014. DOI https://doi.org/10.31789/imscid-2019-001. Disponível em: http://repositorio.pucrs.br/dspace/handle/10923/6953. Acesso em: 16 abril 2018.

FINATTO, M. J. B. O Papel Da Definição De Termos Técnico-Científicos**. Revista da ABRALIN**, v.1, n.1, p. 73-97, julho 2002. DOI https://doi.org/10.5380/rabl.v1i1.52704. Disponível em: http://www.abralin.org/revista/RV1N1/artigo3/RV1N1_art3.pdf. Acesso em: 16 abril 2018.

FONSECA FILHO, C. **História da computação**: O Caminho do Pensamento e da Tecnologia. Porto Alegre: EDIPUCRS, 2007. Disponível em: http://www.pucrs.br/edipucrs/online/historiadacomputacao.pdf. Acesso em: 16 Aabril 2018.

FREITAS, F. de S. O Estado da arte da interpretação automática: do pós-guerra aos *apps* de tradução automática de fala. 2016. 159 f. **Monografia** (Bacharelado em Tradução) - Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2016. DOI https://doi.org/10.14393/19834071.v26.n2.2017.38402

FREITAS, F. de S.; ESQUEDA, M. D. Interpretação automática ou tradução automática de fala: conceitos, definições e arquitetura de software. **Tradterm**, São Paulo, v. 29, Julho/2017, p. 104-145.

FROMM, G. A Construção do Sentido em Vocabulários Técnicos: o Uso de Corpora e Outros procedimentos. **Crop**, São Paulo, v. 10, p. 225-239, 2005. DOI https://doi.org/10.11606/d.6.2011.tde-09092011-160114. Disponível em: http://comet.fflch.usp.br/sites/comet.fflch.usp.br/files/u30/from_tecnico.pdf. Acesso em: 16 abril 2018.

FÜGEN, C. **A system for simultaneous translation of lectures and speeches**. 2008. 204f. Tese (Doutorado). Fakultät für Informatik, Universität Fridericiana zu Karlsruhe, 2008. Disponível em: https://d-nb.info/1014223113/34?origin=publication_detailsam. Acesso em: 10 abril 2018.

GRAZINA, N. M. M. **Automatic Speech Translation**. Dissertação (Mestrado) - Instituto Superior Técnico, Universidade de Lisboa, Lisboa, 2010. DOI https://doi.org/10.32385/rpmgf.v29i2.11056. Disponível em: http://www.inesc-id.pt/pt/indicadores/Ficheiros/5512.pdf. Acesso em: 10 abril 2018.

HASHIMOTO, K. *et al.* Impacts of machine translation and speech synthesis on speech-to-speech translation. **Speech Communication**, v. 54, n.7, p. 857-866, 2012. DOI https://doi.org/10.1016/j.specom.2012.02.004. Disponível em: http://www.sciencedirect.com/science/article/pii/S0167639312000283. Acesso em: 16 abril 2018.

HAUGH, M.; JASZCZOLT K. M. Speaker intentions and intentionality. *In:* **The Cambridge handbook of pragmatics**, 2012, p. 87-112. DOI https://doi.org/10.1017/cbo9781139022453.006. Disponível em: http://people.ds.cam.ac.uk/kmj21/Haugh-Jaszczolt.CUP.Dec10.pdf. Acesso em: 10 abril 2018.

HONG, Y.; KOO, M.; YANG, G. A Korean morphological analyzer for speech translation system. *In*: FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE, 1996, v. 2, p.673-676, 1996, Filadélfia. **Proceedings**… Filadelfia: University of Delaware Alfred I. DuPont Institute, 1996. DOI

https://doi.org/10.1109/icslp.1996.607451. Disponível em: https://ieeexplore.ieee.org/abstract/document/607451/. Acesso em: 10 abril 2018.

IEEE Xplore Digital Library [Internet]. [s.l.]: IEEE. 1998. Disponível em: http://ieeexplore.ieee.org/Xplore/home.jsp. Acesso em: 16 abril 2018.

IIDA, H.; SUMITA E.; FURUSE O. Spoken-language translation method using examples. *In:* THE SIXTEENTH CONFERENCE ON COMPUTATIONAL LINGUISTIS, v. 2, p.1074-1077, 1996, Copenhagen. **Proceedings**... Copenhagen: Association for Computational Linguistics, 1996. DOI https://doi.org/10.3115/993268.993369 Disponível em: http://delivery.acm.org/10.1145/1000000/993369/p1074-iida.pdf?ip=179.104.196.21&id=993369&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1523414324_c5e0579d37a6f0fc60e3aa405d089e7b. Acesso em: 10 abril 2018.

International Center for Advanced Communication Technologies - InterACT [Internet]. Pittsburgh e Karlsruhe: CMU e KIT. 2004 - [citado em 2016 jan. 12]. Disponível em: http://isl.anthropomatik.kit.edu/cmu-kit/english/2162_2673.php. Acesso em: 16 abril 2018.

JEKAT, S.; KLEIN, A. **Machine Interpretation**: Open Problems and Some Solutions. Interpreting, Amsterdam, v. 1, n. 1, p. 7-20, 1996.

KIM, M.; KIM, H. Integrated neural network model for identifying speech acts, predicators, and sentiments of dialogue utterances. **Pattern Recognition Letters,** v. 101, p. 1-5, jan. 2018. DOI https://doi.org/10.1016/j.patrec.2017.11.009. Disponível em: https://www.sciencedirect.com/science/article/pii/S0167865517304129. Acesso em: 10 abril 2018.

KUREMATSU, A. *et al*. Language Processing in connection with Speech Translation at ATR Interpreting Telephony Research Laboratories. **Speech Communication**, v. 10, n. 1, 1991. DOI https://doi.org/10.1016/0167-6393(91)90023-m. Disponível em: https://www.sciencedirect.com/science/article/pii/016763939190023M. Acesso em: 10 abril 2018.

LABOV, W. **Padrões sociolinguísticos**. Trad. Marcos Bagno, Maria Marta Pereira Scherre, Caroline. Rodrigues Cardoso. São Paulo: Parábola, 2008.

LANGLEY, C. Analysis for speech translation using grammar-based parsing and automatic classification. *In:* THE ACL STUDENT RESEARCH WORKSHOP, 2002. **Proceedings**…, 2002. Disponível em: http://www.cs.cmu.edu/~clangley/papers/acl-02-student-research-workshop.pdf. Acesso em: 10 abril 2018.

LEE, T. Speech Translation. *In:* CHAN, S. (org.). **The Routledge Encyclopedia of Translation Technology**. Londres/Nova York: Routledge, 2015, p. 619-631. Disponível em: http://bookzz.org/book/2470011/5925f6. Acesso em: 16 abril 2018.

LEVIN, L. *et al*. An Interlingua Based on Domain Actions for Machine Translation of Task-Oriented Dialogues. *In:* FIFTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, v. 4/7, p. 1155-1158, 1998, Sydney. **Proceedings**… Sydney: Australian Speech Science and Technology Association, 1998. Disponível em: https://www.isca-speech.org/archive/archive_papers/icslp_1998/i98_0999.pdf. Acesso em: 10 abril 2018.

MORIMOTO, T. *et al.* A spoken language translation system: SL-trans2. *In:* THE FIFTEENTH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, v. 2, p. 1048-1052, 1992, Nantes. **Proceedings**… Nantes: Association for Computational Linguistics, 1992. DOI https://doi.org/10.3115/992383.992439. Disponível em: http://www.aclweb.org/anthology/C92-3164. Acesso em 16 abril 2018.

MORIMOTO, T.; KUREMATSU, A. Automatic Speech Translation at ATR. *In:* MT SUMMIT IV, 1993, Kobe. **Proceedings**… Kobe: AAMT, 1993. Disponível em: <http://www.mt-archive.info/MTS-1993-Morimoto.pdf>. Acesso em: 16 abril 2018.

NAKAMURA, S. Overcoming the language barrier with speech translation technology. **Science & Technology Trends**. Tóquio, n.31, abr. 2009. Disponível em: http://www.nistep.go.jp/achiev/ftx/eng/stfc/stt031e/qr31pdf/STTqr3103.pdf. Acesso em: 16 abril 2018.

PAGURA, R. J. **A Interpretação de Conferências no Brasil**: história de sua prática profissional e a formação dos intérpretes brasileiros. 2010. 231f. Tese (Doutorado). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2010. DOI https://doi.org/10.11606/t.8.2010.tde-09022011-151705. Disponível em: http://www.teses.usp.br/teses/disponiveis/8/8147/tde-09022011-151705/publico/2010_ReynaldoJosePagura.pdf. Acesso em: 16 abril 2018.

PINTO, J. H. S. **Um estudo empírico sobre máquinas de tradução em tempo real para equipes distribuídas de desenvolvimento de software**. 2016. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio Grande do Sul. DOI https://doi.org/10.1590/s2176-6681/380213870. Disponível em: http://cbsoft.org/articles/0000/0528/WTDSoft.pdf. Acesso em: 16 abril 2018.

PÖCHHACKER, F. **Introducing Interpreting Studies**. Londres: Routledge, 2004.

SAGISAKA, Y. Spoken Output Technologies. *In:* Survey of the State of the Art in Human Language Technology. *In:* MARIANI, Joseph, *et al.* (org.). **Survey of the State of the Art in Human Language Technology**. 1992. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.7794&rep=rep1&type=pdf. Acesso em: 9 abril 2018.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, jul./oct., p. 379–423, 623–656, 1948. DOI https://doi.org/10.1002/j.1538-7305.1948.tb00917.x. Disponível em: http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf. Acesso em: 20 mar. 2018.

WAIBEL, A. *et al*. Speechalator: two-way speech-to-speech translation on a consumer PDA. *In:* EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, 2003, Genebra. **Proceedings...** Genebra: International Speech Communication Association, 2003. p. 369–372. DOI https://doi.org/10.3115/1073427.1073442. Disponível em: https://www.cs.cmu.edu/~awb/papers/eurospeech2003/speechalator.pdf. Acesso em: 16 abril 2018.

WAIBEL, A.; FÜGEN, C. Spoken language translation - enabling cross-lingual human-human communication. *In:* **IEEE Signal Processing Magazine**, n. 3, 2008. DOI https://doi.org/10.1109/msp.2008.918415. Disponível em: http://isl.anthropomatik.kit.edu/cmu-kit/english/2162_2673.php. Acesso em: 16 abril 2018.

YANG, J.; PARK, J. An experiment on Korean-to-English and Korean-to-Japanese spoken language translation. *In:* IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP-97), 1997, Munique. **Proceedings…** Munique: IEEE, 1997. DOI: https://doi.org/10.1109/icassp.1997.599554. Disponível em: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=599554&newsearch=true&queryText=An%20experiment%20on%20Korean-to-English%20and%20Korean-to-Japanese%20spoken%20language%20translation. Acesso em: 16 abril 2018.

ZONG, C.; SELIGMAN, M. Toward practical spoken language translation. **Machine Translation**, [s.l.], v.19, n.2, p.113-137, 2005. DOI https://doi.org/10.1007/s10590-006-9000-z. Disponível em: http://www.spokentranslation.com/news/pdf/TowardPracticalSpokenLanguageTranslation.pdf. Acesso em: 16 abril 2018.

Artigo recebido em: 21.09.2018 Artigo aprovado em: 19.03.2019