



ENTREVISTA

com

Prof. Dr. Tiago Timponi TORRENT

Por Fernanda Raquel Oliveira LIMA¹



Tiago Timponi Torrent é professor associado do Departamento de Letras e do Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora (UFJF), líder do grupo de pesquisa *FrameNet* Brasil (www2.ufjf.br/framenetbr/), coordenador no laboratório desse grupo, trabalhando no desenvolvimento de recursos lexicais e sintáticos para o português brasileiro. Possui experiência nas áreas de Linguística Cognitiva e Linguística Computacional, com ênfase em Gramática das Construções e Semântica de *Frames*. Doutor em Linguística pela Universidade Federal do Rio de Janeiro (UFRJ), realizou estágio de doutoramento, como *Visiting Student Researcher*, junto ao Departamento de Linguística da University of California, Berkeley. Possui graduação em Letras (Português e Latim) pela Uni-

¹ Doutora em Linguística pela Universidade Federal de Juiz de Fora. Docente do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), *Campus* São Paulo. Endereço eletrônico: <frol@ifsp.edu.br>.



versidade Federal de Juiz de Fora (UFJF) e mestrado em Letras/Linguística pela mesma universidade.

REVISTA METALINGUAGENS – Fernanda Raquel de Oliveira LIMA

Vamos começar falando do trabalho de mais de uma década, coordenado por você, que foi recentemente premiado pela Abralín pelo desenvolvimento de tecnologia e inovação em pesquisas linguísticas (2021) – o Laboratório *FrameNet* Brasil de Linguística Computacional da Universidade Federal de Juiz de Fora. Conte-nos um pouco da história de desenvolvimento desse laboratório e suas filiações.

Tiago Timponi TORRENT

A *FrameNet* Brasil nasceu pelas mãos da Professora Margarida Salomão em 2007. Ela havia retornado de um estágio de pós-doutoramento sênior na University of California, Berkeley, onde reencontrou o Professor Charles Fillmore, que havia sido orientador de doutorado dela na década de 1980 e que é o fundador não só da Semântica de *Frames*, mas também da *FrameNet*. Desse reencontro, nasceu o braço brasileiro da *FrameNet*, a princípio como um projeto de pesquisa, depois, como um grupo de pesquisas, em 2010 e, por fim, como um laboratório, em 2012.

Há dois marcos importantes nesse ano: a constituição do espaço físico da *FrameNet* Brasil e a realização da Escola Internacional de Altos Estudos em Semântica de *Frames* e Aplicações Tecnológicas, ministrada pelos pesquisadores Miriam Petruck e Michael Ellsworth, da Berkeley *FrameNet*. Hoje, em retrospecto, vejo esses dois marcos como importantes sinais de que havíamos alcançado um outro nível de maturidade. Havíamos, pela primeira vez, conseguido sediar uma Escola de Altos Estudos da



CAPES [Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, vinculada ao Ministério da Educação] na UFJF. E havíamos fundado um laboratório de tecnologia numa Faculdade de Letras, contrariando todos os pressupostos.

De 2012 em diante, conseguimos uma série de financiamentos fundamentais, tanto para constituição da infraestrutura física do laboratório quanto para o desenvolvimento tecnológico e para a inserção internacional da nossa pesquisa. Foram mais de 3 milhões de reais captados junto a agências de fomento nacionais, como a FAPEMIG [Fundação de Amparo à Pesquisa do Estado de Minas Gerais], a CAPES e o CNPq [Conselho Nacional de Desenvolvimento Científico e Tecnológico], e internacionais, como a Alexander von Humboldt Foundation e empresas como a Google Inc. Essa verba, bem como – e em especial – tudo aquilo que foi investido pela sociedade brasileira na formação de base dos pesquisadores e discentes do laboratório no âmbito de uma universidade pública, tem gerado resultados muito expressivos. No fim, é isso que faz o laboratório: as pessoas que se formam e trabalham aqui e o retorno que essa formação e esse trabalho trazem para a sociedade.

REVISTA METALINGUAGENS – Fernanda Raquel de Oliveira LIMA

Antes de passarmos para os diferentes desdobramentos das pesquisas dentro do laboratório, você poderia apresentar, rapidamente, a Semântica de *Frames* e explicar o espaço teórico ocupado por ela nesses estudos? Qual seria a importância desse paradigma, entre os outros estudos semânticos, para os trabalhos desenvolvidos dentro do campo da Inteligência Artificial, mais especificamente, no processamento de línguas naturais?



Tiago Timponi TORRENT

A Semântica de *Frames*, conforme definida por Fillmore (1977; 1982; 1985) é uma Semântica da Compreensão. Isso significa dizer que ela se interessa por entender como os usuários de uma língua constroem significados conjuntamente com base no conhecimento que compartilham da situação comunicativa em que se encontram e dos demais aspectos de contexto que sejam relevantes para aquela interação. Nesse sentido, ela se opõe às Semânticas Vericondicionais. O ponto de diferenciação fundamental se manifesta na máxima fillmoreana segundo a qual "[...] os significados são relativizados às cenas" (Fillmore, 1977). Isso significa dizer que, para além da definição de um item lexical, por exemplo, o significado dele também envolve uma cena de fundo que é evocada para sua compreensão. Fillmore exemplifica essa questão opondo as palavras *land* e *ground* em inglês. Enquanto a primeira indica uma porção de terra considerada sob o ponto de vista da sua relação com o mar, a segunda o faz em relação ao ar. Eu, particularmente, gosto muito de usar o exemplo da relação entre as palavras recreio e intervalo. Do ponto de vista definicional, ambas poderiam ser consideradas sinônimas, visto que indicam uma porção de tempo entre duas aulas destinada ao descanso e à alimentação. Entretanto, se eu disser aos meus alunos de graduação que eles estão liberados para ir para o recreio, eles vão, minimamente, achar estranho. Isso porque recreio evoca uma cena de fundo em que o papel do estudante é ocupado por pessoas de uma certa faixa etária. Porque os *frames* também podem codificar diferentes perspectivas sobre uma mesma cena. O modelo também permite dar conta de casos como o de carcereiro *versus* prisioneiro, em português. Uma análise composicional estaria fadada a concluir que trata-se de sinônimos, uma vez que prisão e cárcere o são e



que o sufixo aplicado, -eiro, é o mesmo. Porém, cada uma dessas formações lexicais coloca em proeminência um dos elementos do *frame* de Prisão: enquanto carcereiro focaliza o Operador, prisioneiro focaliza o Interno.

Nesse contexto, como a Semântica de *Frames* é capaz de capturar aspectos fundamentais do significado dos elementos linguísticos, em especial aqueles que estão ancorados no senso comum, ela tem um grande potencial para aplicações em Inteligência Artificial. A Compreensão de Língua Natural por máquinas requer o acesso a representações de conhecimentos para além das propriedades formais dos elementos linguísticos e também para além de uma semântica composicional.

REVISTA METALINGUAGENS – Fernanda Raquel de Oliveira LIMA

De que maneira essa teoria, implementada na forma de modelos cognitivamente inspirados para compreensão de língua natural, pode melhorar o desempenho de tradutores automáticos?

Tiago Timponi TORRENT

A tradução automática é a ponta de lança da Inteligência Artificial. Ao longo da história, sempre a tradução automática experimentou grandes avanços, o mesmo ocorreu com a IA. O primeiro grande experimento em tradução automática, conduzido pela IBM na Georgetown University, em 1954, ocorreu apenas dois anos antes da primeira conferência em IA. O inverno da IA, entre a década de 1970 e o início dos anos 2000, também coincidiu com uma certa estagnação da tradução automática.

Dos anos 2000 para cá, com o acesso a imensos volumes de dados linguísticos disponíveis eletronicamente, a tradução automática encontrou nos modelos matemáticos – como as redes neurais artificiais – uma impor-



tante alavanca para seu desenvolvimento. Modelos como o *OpenNMT*, o *Google Translate* e o *DeepL* usam redes neurais treinadas em grandes corpora bilíngues. Tais sistemas, como toda rede neural, são muito bons em abstrair padrões. Assim, pela manipulação recursiva das formas linguísticas, eles chegam a resultados muito bons. Porém, como nós linguistas cognitivistas que lemos Fauconnier e Turner (2002) sabemos, a língua é só a ponta do *iceberg*. Os processos de significação são iniciados pela forma linguística, que convida o interlocutor a co-construir com o enunciador o significado de um enunciado qualquer. Essa co-construção é baseada não só naquilo que se diz, no texto em si, mas também em conhecimento de mundo compartilhado. Assim, é quase impossível que a mera manipulação do pólo formal da linguagem resulte em sistemas que repliquem a capacidade humana.

Com essa bagagem teórica em mente, desenvolvemos na *FrameNet* Brasil um algoritmo para tentar servir de prova de conceito para a ideia de que um modelo de língua cognitivamente inspirado pode contribuir para melhorar o desempenho de tradutores automáticos, em especial naqueles contextos em que o conhecimento de mundo sobre um domínio específico é necessário. Assim foi criado o *Scylla* (Costa et al. 2022): um *pipeline* de pós-edição para tradutores automáticos baseados em redes neurais que usa a *FrameNet* Brasil como um modelo contextualizado do domínio dos Esportes. O projeto é baseado na tese de doutorado do Alexandre Diniz da Costa, que modelou o domínio dos Esportes Olímpicos na *FrameNet* Brasil por meio da criação de cerca de 40 novos *frames*, centenas de unidades lexicais e milhares de relações entre os *frames* e os itens lexicais. Tudo isso em duas línguas: português e inglês.



O *Scylla* funciona da seguinte forma: dada uma sentença numa língua fonte, por exemplo, o português brasileiro, ele submete essa sentença a um tradutor neural qualquer – no caso, usamos a API do *Google Translate* – para que ela seja traduzida para a língua alvo, no nosso exemplo, o inglês. No experimento que conduzimos, uma das sentenças era "O ponta é o jogador que menos tempo tem para pensar na armação de uma jogada" e a tradução que a API do *Google Translate* gerou para ela foi "*The forward is the player who has less time to think about setting up a move*". A princípio, a tradução parece boa. Porém, ela tem uma inadequação no contexto do futebol. A melhor tradução para *ponta.n* não é *forward.n*, mas *winger.n*, já que *winger.n* é um subtipo de *forward.n*, assim como *ponta.n* é um subtipo de *atacante.n*. O tradutor neural não capturou essa questão, possivelmente por questões relativas ao *corpus* de treinamento ou aos padrões de associação de formas linguísticas aprendidos no treinamento.

Mas, no *Scylla*, essas duas sentenças são submetidas a um outro software criado pela *FrameNet* Brasil chamado *DAISY*, que analisa automaticamente os *frames* evocados por cada palavra em uma sentença a partir do contexto da sentença em si. A *DAISY* então, analisando o contexto, vê que *ponta.n* é uma unidade lexical do *frame* de *Atletas_por_posição* e que, na *FrameNet* Brasil, a melhor tradução para essa palavra é *winger.n*. Então, ao final do *pipeline*, ela substitui *forward.n* por *winger.n* e "corrige" a sentença gerada pela API do Google Tradutor para "*The winger is the player who has less time to think about setting up a play*". Nós avaliamos a performance do *Scylla* em contraste com a do *Google Translate* para 50 pares de sentenças do domínio do Esportes e o *Scylla* melhora a qualidade da tradução em 47% para uma métrica chamada HTER, que mede o esfor-



ção de edição que um tradutor humano teria que fazer para adequar uma tradução automática.

REVISTA METALINGUAGENS – Fernanda Raquel de Oliveira LIMA

Ao longo desses anos de pesquisa no Laboratório *FrameNet* Brasil, já foram desenvolvidas significativas aplicações das bases de dados, como um dicionário para a Copa do Mundo no Brasil, um assistente de viagem, um aplicativo para criação de *frames* nos celulares, um tradutor automático e um sistema de recomendação de atrações turísticas baseado em uma interface conversacional, dentre outras. Seria possível fazer uma apresentação ampla desses ganhos de pesquisa?

Tiago Timponi TORRENT

Fazer pesquisa aplicada é, ao mesmo tempo, um desafio e um estímulo. É desafiador porque o problema de pesquisa tem um impacto mais imediato na sociedade, o que faz com que os resultados sejam analisados segundo métricas mais diretas e concretas. É um estímulo porque muitas vezes, em especial para quem está começando na pesquisa, é difícil ver o resultado do esforço hercúleo que se faz durante um mestrado ou um doutorado de forma mais clara. Isso acaba alimentando aquela ideia da síndrome de impostor. Quando o trabalho de um aluno vira um *software* ou uma patente, essa concretude traz um certo alento na trajetória dele.

Desde o Dicionário *FrameNet* Brasil da Copa do Mundo, que foi um produto mais lexicográfico, até os dois depósitos de patentes resultantes do *Scylla* e o *Lutma*, que é uma ferramenta colaborativa para criação de *frames*, o que eu observo é que pensar no produto tecnológico final e vê-lo tomar forma também nos faz repensar o próprio trajeto da pesquisa. É



como se as questões que emergem da confecção do produto tecnológico retroalimentassem o processo de construção teórico-metodológica do objeto de análise. Parafraseando o pai da Linguística, é como se o fato de haver um produto tecnológico sendo desenvolvido como resultado daquela pesquisa acabasse por reenquadrar o ponto-de-vista e, por consequência, o objeto por ele criado.

REVISTA METALINGUAGENS – Fernanda Raquel de Oliveira LIMA

No curso de Letras da UFJF há disciplinas que não são comumente encontradas nesse mesmo curso em outras instituições de ensino – como a oficina de práticas de anotação mediada por computador, tópicos em estudos cognitivos da gramática e do léxico e tópicos em Linguística Computacional. De que maneira o senhor tem observado a importância desses estudos já na graduação para a formação do profissional egresso das Letras?

Tiago Timponi TORRENT

A profissão do futuro para o Linguista já existe e atende pelos nomes de *Language Engineer*, *Language Research Scientist*, *Language Specialist*, entre outros. Empresas como Google, Meta, Amazon, Microsoft e demais gigantes tecnológicas já sabem há mais de uma década que o futuro da interação humano-máquina passa pelas línguas naturais e que os linguistas são parte fundamental nesse processo. Essas disciplinas buscam aproximar os graduandos dessas novas oportunidades de carreira na indústria tecnológica, tanto pelo viés mais teórico – por exemplo, nos Tópicos em Linguística Computacional –, quanto pelo mais prático – na Oficina de Linguagem e Tecnologia. Na Oficina, a gente sempre trabalha com algum projeto de anotação real, para que os discentes possam ter uma vi-



são holística de como é trabalhar com Linguística Computacional. Sempre que possível, envolvemos parceiros externos no processo e socializamos os resultados da anotação com os alunos.

Já os Tópicos em Linguagem e Cognição são o marco da tradição fundada na UFJF pela Margarida Salomão, quando ela retornou do doutorado em Linguística da University of California, Berkeley na década de 1990. Linguística Cognitiva sempre foi um diferencial da UFJF e esse enquadramento cognitivista está também na forma como nós fazemos Tecnologia Linguística na *FrameNet* Brasil.

REVISTA METALINGUAGENS – Fernanda Raquel de Oliveira LIMA

Aproveitando o gancho da pergunta anterior, de que formas os estudos desenvolvidos no Laboratório *FrameNet* Brasil de Linguística Computacional integram alunos da graduação e da pós-graduação? Há projetos envolvendo outras instituições ou mesmo outras áreas do conhecimento, parcerias para além das academias?

Tiago Timponi TORRENT

Ao longo da última década, mais de 60 alunos se vincularam ao laboratório. Nesse grupo, tivemos alunos de iniciação científica, mestrandos, doutorandos e pós-doutorandos, tanto vinculados ao curso de Letras e ao PPG-Linguística, quanto aos cursos de Ciência da Computação, Sistemas de Informação, Design e ao PPG-Ciência da Computação da UFJF. Essa interdisciplinaridade é uma marca do laboratório e algo de que não abrimos mão. Nosso corpo técnico, para além dos professores-pesquisadores dos departamentos de Letras e Ciência da Computação, conta com um analista de TI e um jornalista.



Para além dessa configuração institucional, a *FrameNet* Brasil ainda mantém um conjunto de parcerias com universidades e instituições de pesquisa no Brasil e no exterior. Recentemente, aprovamos junto à FAPESP a criação da ReINVenTA – *Research and Innovation Network for Visual and Textual Analysis of Multimodal Objects* –, uma rede de pesquisa coordenada pela *FrameNet* Brasil que congrega ainda o Laboratório Experimental de Tradução (LETra) da UFMG, além de parceiros na UFU e na PUC-MG. A ReINVenTA receberá um aporte de mais de 1 milhão de reais para o desenvolvimento de modelos de inteligência artificial para o processamento computacional de gêneros multimodais, tais como programas de TV, por exemplo.

No âmbito internacional, destacam-se três importantes parcerias consolidadas. Em primeiro lugar, temos a parceria com a matriz da *FrameNet* no International Computer Science Institute, em Berkeley, Califórnia. Desde 2007, nossos intercâmbios com o ICSI têm se mantido ativos. Já recebemos pesquisadores, enviamos pesquisadores e alunos e atuamos juntos em diversos projetos de pesquisa. Desde 2015, temos também uma parceria com o Departamento de Sueco, Multilinguagem e Tecnologia Linguística da Göteborgs Universitet, na Suécia. Essa parceria, financiada pela CAPES e pelo STINT, possibilita o desenvolvimento de pesquisa conjunta na área de modelagem computacional de construções gramaticais, bem como o intercâmbio de pesquisadores e discentes, recorrendo a missões de trabalho e bolsas de doutorado sanduíche e pós-doutorado. Em 2018, firmamos uma nova parceria, financiada pela CAPES e pelo DAAD, com duas universidades na Alemanha: a Heinrich-Heine Universität Düsseldorf e a Leipzig Universität. Nessa colaboração, também realizamos intercâmbios anuais via missões de trabalho e bolsas de doutorado e pós-dou-



torado no exterior. O foco, porém, é no alinhamento de *lexicons* e *constructicons* computacionais para melhorar o desempenho de tradutores automáticos.

Temos também uma parceria muito importante como o Distributed Little Red Hen Lab, coordenado pelos professores Mark Turner e Francis Steen. Foi essa parceria que possibilitou o desenvolvimento do *Lutma* (Torrent *et al.*, 2022), nosso software para criação colaborativa de *frames*, que foi financiado pela Alexander von Humboldt Foundation, por meio de um *Anneliese Meier Research Award* concedido ao Turner.

A parceria com o Red Hen também se mantém desde 2019 por intermédio de projetos conjuntos submetidos ao programa de desenvolvimento de inovação *Google Summer of Code*. Por esse programa, a *Google* paga bolsas de desenvolvimento tecnológico a estudantes de TI do mundo inteiro que se candidatam a desenvolver projetos sob a mentoria de laboratórios credenciados. Tanto a *FrameNet* Brasil quanto o Red Hen participam desse programa. Ao longo dos anos, já recebemos estudantes da Índia, Escócia, Rússia e Estados Unidos.

Por fim, mas não menos importante, não poderia deixar de mencionar a *Global FrameNet*, iniciativa internacional de pesquisa e desenvolvimento cuja criação foi liderada por nós e que congrega mais de uma dezena de laboratórios em instituições de ensino e pesquisa espalhadas pelas Américas, Europa e Ásia.

Envio: Setembro de 2022
Aceite: Dezembro de 2023.